

¿Leen los Universitarios las Encuestas de Evaluación del Profesorado Cuando se Aplican Incentivos por Participación? Una Aproximación Empírica

Do Students Read Teacher Evaluation Surveys when Participation Incentives are Applied? An Empirical Approach

Luis Matosas-López *
Alberto Romero-Ania
Elena Cuevas-Molano

Universidad Rey Juan Carlos, España

El objetivo del estudio es revelar en qué medida los universitarios leen las encuestas de evaluación del profesorado cuando se aplican incentivos por participación. Se desarrolla un estudio de carácter cuantitativo, en el que, se adopta una metodología de tipo experimental con dos grupos. El primero realiza la valoración de su docente en un escenario libre de incentivos, el segundo completa la encuesta en un escenario de participación incentivada. El estudio considera además dos tipologías de cuestionario: por un lado, escalas de Likert y, por otro, escalas con episodios de comportamiento BARS. La investigación emplea análisis descriptivos, pruebas t-Student y análisis de correlaciones de Pearson. Los resultados revelan diferencias en el tiempo invertido cuando la participación es incentivada. Se concluye que los instrumentos con escalas de Likert no parecen favorecer una óptima lectura y cumplimentación de la encuesta cuando la evaluación introduce recompensas. Esta situación puede ser mejorada empleando cuestionarios BARS. El presente estudio arroja luz sobre un problema prácticamente ignorado por investigaciones previas, sino que, además, aborda el mismo proponiendo alternativas de mejora.

Descriptores: Calidad de la educación; Evaluación del docente; Eficacia del Docente; Universidad; Cuestionario.

The purpose of this study is to reveal the extent to which university read teacher evaluation surveys when participation incentives are applied. Researchers carry out a quantitative study, in which an experimental methodology with two groups is adopted. The first group performs the assessment of the teacher in a scenario free of incentives; the second completes the survey in an incentivized participation scenario. In addition, the research considers two types of questionnaires: on the one hand, Likert scales, on the other hand, scales with behavioral episodes or BARS. The research uses descriptive analysis, Student's t-test, and analysis of correlations through the Pearson correlation coefficient. The findings reveal differences in the investment of time when participation incentives are applied. It can be concluded that the instruments with Likert scales do not favor the correct reading and completion of surveys when the evaluation introduces rewards. However, this situation can be improved using BARS questionnaires. The present study sheds light on a problem practically ignored by previous literature, but also introduces alternatives for improvement.

Keywords: Educational quality; Teacher evaluation; Teacher effectiveness; Universities; Questionnaires.

*Contacto: luis.matosas@urjc.es

Introducción

La evaluación de la calidad docente en la educación universitaria ha sido objeto de extensas investigaciones. Aunque son muchos los autores que presentan serias reservas a la hora de medir la calidad docente mediante encuestas de evaluación del profesorado (Galbraith y Merrill, 2012; Moreno Olivos, 2018; Reyero, 2014), sus resultados están ampliamente aceptados como indicador de eficiencia. Tanto es así que la visión plasmada por el estudiante sobre el desempeño del profesor en estos cuestionarios está, desde hace décadas, considerada un elemento clave a la hora de determinar la calidad de la docencia en universidades de todo el mundo (Darwin, 2017; De-Juanas Oliva y Beltrán Llera, 2013; Guzmán, 2018; Ruiz Carrascosa, 2000; Tejedor Tejedor, 2009).

Las encuestas de evaluación del profesorado pueden adoptar distintos formatos. Estos van desde cuestionarios con preguntas abiertas o cuestionarios que combinan preguntas abiertas con preguntas cerradas (Nasser-Abu y Fresko, 2009), a cuestionarios con episodios de comportamiento o BARS (Behaviorally Anchored Rating Scales) (Martin-Raugh et al., 2016; Matosas-López y Leguey-Galán, 2018). Sin embargo, la mayor parte de las instituciones universitarias optan por cuestionarios con escalas Likert, en los que el alumno manifiesta su nivel de acuerdo con uno o varios ítems descriptivos de la categoría a evaluar (Lizasoain, Etxeberria y Lukas, 2017; Molero y Ruiz, 2005; Muñoz, Ríos y Abalde, 2002).

Sea cual sea el formato adoptado en el cuestionario, las encuestas de evaluación del profesorado están sujetas a constante debate. Entre los temas objeto de discusión destacan: la dimensionalidad de las mediciones, la fiabilidad y la validez de los instrumentos de medición, el error de lenidad, el efecto halo, las variables de sesgo que afectan al resultado o las tasas de participación del alumnado.

El número de dimensiones a contemplar en la labor docente o dimensionalidad es uno de los temas controvertidos. El SEEQ (Student Evaluations of Educational Quality), instrumento originalmente diseñado por Marsh (1982) y uno de los sistemas de medición más utilizados, presenta un modelo con nueve dimensiones: valor de aprendizaje, entusiasmo, organización, interacción grupal, relación individual, cobertura, exámenes o evaluación, tareas y carga de trabajo. Marsh (1991) afirma que los ítems de su cuestionario pueden ser aglutinados mediante análisis factorial en grupos de ítems correlacionados entre sí o dimensiones, y que, en consecuencia, estas dimensiones han de ser interpretadas de manera independiente. Por el contrario, Abrami y d'Apollonia (1997) afirman que la calidad docente puede ser medida con una única dimensión o factor global denominado "habilidad de instrucción general". Este factor global deriva de la combinación de tres habilidades subyacentes: impartir formación, facilitar las interacciones y evaluar el progreso del alumno.

Muchos son los trabajos que analizan la fiabilidad, la validez, el error de lenidad y el efecto halo en estas encuestas. Diversas investigaciones revelan que las valoraciones realizadas por los estudiantes sobre la labor del profesor pueden ser consideradas una medida fiable y razonablemente válida a la hora de evaluar la calidad docente (Feistauer y Richter, 2016; Spooren, Mortelmans y Christiaens, 2014). Sin embargo, algunos autores muestran reticencias en cuanto al uso de indicadores de consistencia interna para hacer atribuciones sobre la fiabilidad de los instrumentos de medición en sí mismos (Marsh, 1987; Morley, 2012). De igual forma, el error de lenidad o el efecto halo son otros de los aspectos que

ejercen una influencia directa en estas encuestas. El error de lenidad se define como una alteración en las puntuaciones promedio en una dirección determinada de la escala (Sharon y Bartlett, 1969), mientras que el efecto halo describe la tendencia de los encuestados a colocar al docente calificado al mismo nivel en varias de las categorías sujetas a valoración (Bernardin, 1977).

La literatura de este campo presta también gran atención a las variables de sesgo exógenas a la medición que ejercen algún tipo de influencia en esta. Entre estas variables de sesgo, la lenidad de las calificaciones es materia de frecuente discusión. Diferentes autores reportan correlaciones entre las calificaciones esperadas por el alumno y la valoración que este hace de su profesor (Griffin, 2004; McPherson, 2006). El tipo de asignatura (opcional vs obligatoria) o la disciplina del curso (artes o humanidades vs ciencias) están también correlacionados con las evaluaciones docentes. Feldman (1978) afirma que las valoraciones del profesorado tienden a ser más altas en asignaturas optativas y asignaturas de artes o humanidades que en materias obligatorias o del ámbito de las ciencias. Asimismo, el tamaño del grupo (Gannaway, Green y Mertova, 2017), el rango académico del docente (Spooren, 2010), el carácter del alumno (McCann y Gardner, 2014), o el género (Boring, 2017) son otros ejemplos habituales de sesgos correlacionados con las encuestas de evaluación del profesorado.

Otra de las cuestiones objeto de controversia es la baja tasa de participación entre el alumnado. Esta situación se ha visto además acentuada con la implementación de encuestas que usan formularios *online*. Son varios los motivos que ocasionan esta baja participación del estudiante, pero quizás el más determinante es la falta de anonimato percibida por el mismo a la hora de realizar la evaluación (Sorenson y Reiner, 2003). Este recelo en torno a la confidencialidad de las respuestas se genera cuando el alumno ha de emplear sus credenciales para acceder al formulario de recogida de datos. Las ventajas de los sistemas de administración en línea son incontestables. Los cuestionarios *online* eliminan los costes de impresión, distribución, recolección, escaneado y transcripción de datos (Dommeyer et al., 2004; Nair y Adams, 2009); sin embargo, son muchos los estudios que constatan que las tasas de participación obtenidas con estos cuestionarios son muy inferiores a las alcanzadas en clase con cuestionarios en papel (Nulty, 2008; Stanny y Arruda, 2017; Stowell, Addison y Smith, 2012).

1. Lectura y cumplimentación de la encuesta

Si bien las investigaciones previas han estudiado extensamente temas como los ya señalados, la literatura del acervo científico deja sin abordar un aspecto clave en estas mediciones, y que hasta el momento ha resultado prácticamente tabú entre la comunidad universitaria: ¿realmente leen los estudiantes las encuestas de evaluación del profesorado cuando se aplican incentivos por participación? La revisión de los trabajos, sobre valoración docente, publicados desde 2009 en revistas indexadas en los catálogos Web of Science (Clarivate Analytics) y Scopus (Elsevier) revela la usencia total de estudios que aborden específicamente esta temática.

Esta cuestión es, además, un punto que reviste especial importancia en el momento actual. La implementación de sistemas de valoración *online* en todo el mundo y las bajas tasas de participación experimentadas en ellos han llevado a las instituciones universitarias a adoptar estrategias de mejora de la participación que no parecen favorecer este aspecto.

Entre las estrategias empleadas destacan aquellas que recurren al uso de incentivos como forma de recompensa para aquellos estudiantes que cumplimentan la encuesta. En este capítulo encontramos beneficios de muy distinta índole: acceso anticipado a las calificaciones, créditos adicionales, puntos extra, e incluso beneficios dentro de los servicios de la facultad como puedan ser regalos de librería o cupones de cafetería (Ballantyne, 2003; Dommeyer et al., 2004; Johnson, 2003). La utilización de estos incentivos puede mejorar la participación, sí, pero en opinión de los autores conduce también a que el alumno, con el mero propósito de obtener la recompensa, cumplimente el cuestionario de manera mecánica y acelerada. Por consiguiente, la utilización de incentivos puede resultar contraproducente, llegando a comprometer la correcta lectura y cumplimentación de la encuesta.

No obstante, evaluar si los estudiantes leen realmente las encuestas de valoración docente reviste cierta complejidad. Para este propósito, la alternativa más extendida pasa por la utilización de preguntas de control. Este sistema recurre a la fiscalización de la coherencia entre las respuestas como medio para discernir si el alumno realiza una correcta lectura del cuestionario. Sin embargo, el empleo de preguntas de control conlleva también un aumento sustancial en la longitud de la encuesta, además de una gestión añadida en el posterior tratamiento de las respuestas.

A falta de referencias previas de trabajos que exploren específicamente esta temática, los autores optan por tomar el tiempo invertido en la realización de la encuesta como indicador de la óptima lectura y cumplimentación de esta. Además, la medición del tiempo de realización como medida de control resulta imperceptible para el estudiante y no interfiere de forma alguna en la cumplimentación del cuestionario.

Con el transcurso de los años y el desarrollo tecnológico, los instrumentos de medición *online* han evolucionado incorporando gran variedad de funcionalidades. Entre estas funcionalidades destaca la medición del tiempo invertido por el encuestado. Esta característica posibilita, no solo el control del tiempo desde el momento de apertura del formulario hasta el momento de finalización de este sino también el tiempo empleado en cada pregunta del cuestionario.

Objetivos

Considerando la exposición previa, el presente estudio pretende contestar la pregunta planteada, revelando en qué medida los estudiantes universitarios leen las encuestas de evaluación del profesorado cuando se aplican incentivos por participación. Para ello los autores postulan la inversión de tiempo realizada por el alumno como indicador de la efectiva lectura y cumplimentación de la encuesta. Además, la investigación considera dos tipologías de cuestionario: por un lado, un cuestionario con escalas de Likert y por otro, un cuestionario con episodios de comportamiento BARS, con la intención de identificar posibles diferencias en la inversión de tiempo realizada por el estudiante ante ambos formatos de encuesta.

En consecuencia, los investigadores para dar respuesta a la pregunta ¿leen los universitarios las encuestas de evaluación del profesorado cuando se aplican incentivos por participación? plantean de forma paralela dos objetivos que quedan desglosados de la siguiente manera:

- Determinar la inversión de tiempo experimentada con un cuestionario con escalas Likert incentivado.

- Determinar la inversión de tiempo experimentada con un cuestionario con escalas BARS incentivado.

2. Método

Los investigadores desarrollan un estudio de carácter cuantitativo, en el que, siguiendo las directrices de Hernández Pina (1997), se adopta una metodología de tipo experimental. En el diseño experimental del trabajo cabe detallar la información relativa a los instrumentos empleados, la muestra seleccionada y el procedimiento de recogida y análisis de datos.

Instrumentos

En el estudio se emplean dos instrumentos de medición de la calidad docente diferenciados (estos quedan recogidos en los anexos I y II). Por un lado, un cuestionario con escalas de Likert de cinco grados de 1-Muy en desacuerdo a 5-Muy de acuerdo y, por otro, un cuestionario con episodios de comportamiento BARS con cinco escenarios de comportamiento. Ambos instrumentos contemplaron la evaluación de diez categorías de la labor del docente, siendo cada una de ellas valorada con un único ítem. Los ítems considerados fueron: introducción a la asignatura, descripción del sistema de evaluación, gestión del tiempo, disponibilidad general, coherencia organizativa, implementación del sistema de evaluación, resolución de dudas, capacidad explicativa, facilidad de seguimiento y satisfacción general.

Para garantizar una adecuada comprensión de las preguntas de los cuestionarios, estos fueron evaluados por un grupo de 12 jueces expertos en materia de evaluación educativa (Escobar-Pérez y Cuervo-Martínez, 2008). Asimismo, ambos cuestionarios fueron sometidos a un análisis de validez y fiabilidad. Para el análisis de validez y fiabilidad se contó con la participación de 393 estudiantes de la Facultad de Ciencias Jurídicas y Sociales (FCJS) de la Universidad Rey Juan Carlos (URJC).

En línea con las recomendaciones realizadas por distintos autores en estudios previos, la validez y fiabilidad de ambos instrumentos –tanto Likert como BARS– se examina empleando, por un lado, la técnica de análisis factorial exploratorio (AFE) y, por otro, el coeficiente alfa de Cronbach (Cañadas y Cuétara, 2018; Fernández Millán y Fernández Navas, 2013; Luna Serrano, 2015; Stoskopf et al., 1992).

En ambos instrumentos, el AFE reveló la existencia de dos constructos subyacentes (Aptitud y actitud docente, Estructuración y evaluación) capaces de explicar: el 68,52% de la varianza en el caso del cuestionario Likert y el 65,74% de la varianza en el cuestionario BARS. La fiabilidad, expresada a través del estadístico alfa de Cronbach, mostró un coeficiente de ,94 para el cuestionario tipo Likert y ,93 para el cuestionario BARS. Los porcentajes de varianza total explicada así como los coeficientes alfa de Cronbach en ambos instrumentos respaldan la validez y fiabilidad de estos de cara a su aplicación durante en el estudio (George y Mallery, 2003).

Muestra

La población objeto de estudio quedó constituida por estudiantes de la FCJS de la URJC. Los investigadores, recurriendo a un muestreo intencional o de conveniencia (Mayorga Fernández y Ruiz Baeza, 2002), seleccionaron 8 de los 23 grados de la FCJS. Los grados considerados fueron los siguientes: Grado en Administración y Dirección de Empresas,

Grado Administración y Dirección de Empresas y Derecho, Grado en Marketing, Grado en Criminología, Grado en Relaciones Internacionales, Grado en Economía, Grado en Protocolo y Organización de Eventos y Grado en Trabajo Social. Conforme a la información publicada por la URJC el número total de estudiantes matriculados en estos grados en el curso en el que se realiza el estudio asciende a 9.701.

La investigación cuenta con una muestra heterogénea de 451 estudiantes. Para un nivel de confianza del 95%, y bajo un supuesto de máxima indeterminación con $P = Q = 50\%$ en la población, se trabaja con un error muestral del 4,51%. Dado que es habitual admitir errores muestrales de incluso el 5% (Ficapal-Cusí et al., 2013), el margen de error considerado otorga al estudio una significatividad estadística adecuada.

A la vista de la variedad de titulaciones contempladas, los autores emplearon la técnica de muestreo estratificado con el objeto de mantener la representatividad de los grados en la muestra (Buendía, 1997). La distribución de cada estrato queda recogida en el cuadro 1.

Cuadro 1. Distribución de la población y la muestra

GRADO	POBLACIÓN	MUESTRA	%
Grado en Administración y Dirección de Empresas	3.962	188	4,75%
Grado Administración y Dirección de Empresas y	1.391	63	4,55%
Grado en Marketing	1.801	82	4,55%
Grado en Criminología	763	36	4,45%
Grado en Relaciones Internacionales	706	32	4,70%
Grado en Economía	652	30	4,60%
Grado en Protocolo y Organización de Eventos	284	13	4,50%
Grado en Trabajo Social	142	6	4,65%
<i>Total</i>	<i>9.701</i>	<i>451</i>	<i>4,55%</i>

Fuente: Elaboración propia.

Los 451 estudiantes participantes en el estudio se dividieron en dos grupos, denominados grupo 1 y grupo 2. Los estudiantes del grupo 1 ($n = 229$) abordaron la valoración de sus docentes empleando un cuestionario libre de incentivos. Por su parte, los estudiantes del grupo 2 ($n = 222$) realizaron la evaluación de sus profesores siguiendo el protocolo de valoración docente con aplicación de incentivos empleado habitualmente por la URJC. Tanto la asignación de los participantes a los grupos como la elección de los profesores y materias a evaluar, en línea con Buendía (1994), se realiza de manera incidental seleccionando las alternativas más accesibles en cada caso.

El protocolo de valoración docente con aplicación de incentivos empleado en el grupo 2 persigue maximizar la participación del alumnado, ofreciendo como recompensa la posibilidad de acceder de manera anticipada e inmediata a la calificación final de la asignatura una vez cumplimentado el cuestionario.

A su vez los participantes de cada grupo fueron separados en dos colectivos diferenciados, al primer colectivo se le administró un cuestionario Likert, al segundo un cuestionario BARS. De los 229 estudiantes encuadrados en el grupo 1, 112 efectuaron la valoración empleando un cuestionario Likert, mientras que 117 abordaron la evaluación utilizando un cuestionario BARS. De los 222 participantes ubicados en el grupo 2, 106 realizaron la evaluación a través de un cuestionario Likert y 114 emplearon un cuestionario BARS.

Recogida y análisis de datos

A cada colectivo le fue cargado en su Aula Virtual el tipo de cuestionario correspondiente dentro de la asignatura cuyo profesor iba a ser evaluado. Los formularios quedaron habilitados en el Aula Virtual del alumno durante un periodo de un mes para que el estudiante pudiera abordar la evaluación del docente en remoto. Todos los cuestionarios conservaron el anonimato de los participantes.

Una vez recogidos los datos se procedió al estudio de estos empleando el software IBM SPSS Statistics 25.0.0®. Por medio de este paquete, los datos fueron sometidos a tres tipos de análisis: a) análisis descriptivos, b) análisis de pruebas paramétricas para muestras independientes aplicando el estadístico t-Student, y c) análisis de correlaciones bivariadas empleando el coeficiente de correlación de Pearson. Los dos últimos análisis fueron precedidos de la realización de las pertinentes pruebas de normalidad.

El análisis descriptivo explora los datos de tiempo invertido en la lectura y cumplimentación de la encuesta. Por su parte, el análisis de pruebas paramétricas para muestras independientes tiene el objetivo de revelar si el tiempo de cumplimentación difiere de forma significativa en función del grupo. Por último, el análisis de correlaciones bivariadas explora las correlaciones entre extensión del enunciado y el tiempo promedio invertido en la lectura de cada ítem del cuestionario. Todos los análisis se desarrollan para los grupos 1 y 2, cotejando en cada grupo los resultados para ambas tipologías de cuestionario.

3. Resultados

En el grupo en el que no existe recompensa alguna por completar el cuestionario (Grupo 1) la tasa de participación es del 72,20%. Por su parte en el grupo en que se incentiva la realización de la encuesta (Grupo 2) la tasa de participación alcanza el 84,35%.

En el grupo 1, se mide el tiempo de realización del cuestionario cuando no se ofrece incentivo alguno por su cumplimentación. En el grupo 2, se mide el tiempo que invierte el alumno en la realización del cuestionario en un escenario de incentivo. En primera instancia, en ambos casos se proporcionan los datos descriptivos (M y DT) del tiempo empleado en la cumplimentación del conjunto de la encuesta y de cada una de las diez preguntas para ambos formatos de cuestionario.

Antes de proceder con los análisis siguientes, se realiza una prueba de normalidad sobre las variables extensión del enunciado y tiempo promedio invertido para cada ítem en ambos grupos para las dos tipologías de cuestionario observadas. Para ello, se comprueba que el indicador de normalidad Shapiro-Wilk tiene en todas las variables un p-valor superior a 0,01 (Extensión enunciado Likert = 0,63; M Tiempo Likert Grupo 1 = 0,63; M Tiempo Likert Grupo 2 = 0,93; Extensión enunciado BARS = 0,39; M Tiempo BARS Grupo 1 = 0,65; M Tiempo BARS Grupo 2 = 0,27).

Una vez comprobado que las variables indicadas siguen una distribución normal, se aborda el análisis de pruebas paramétricas para muestras independientes aplicando el estadístico t-Student y el análisis de correlaciones bivariadas mediante el coeficiente de correlación de Pearson. Manteniendo la organización por objetivos planteada en la introducción, para facilitar la interpretación de los hallazgos realizados, así como la comparación de los tiempos de cumplimentación en cada grupo para ambos formatos de encuesta, los

resultados son presentados por tipología de cuestionario. Por un lado, los hallazgos obtenidos empleando el cuestionario Likert, por otro, los conseguidos con el cuestionario BARS.

3.1. Inversión de tiempo experimentada con un cuestionario con escalas Likert incentivado

Los resultados muestran que el tiempo promedio empleado por los participantes en completar el cuestionario Likert ante una situación de incentivo (65,45 segundos) es muy inferior al tiempo promedio para cumplimentar este tipo de cuestionario cuando no media recompensa (111,41 segundos). De igual forma, el tiempo promedio invertido por los participantes del grupo 2 en cada pregunta es inferior al promedio del grupo 1 en todos los ítems (cuadro 2).

Cuadro 2. Inversión de tiempo empleado en un cuestionario Likert en ambos grupos

	Extensión enunciado ^b	GRUPO 1 ^a		GRUPO 2 ^a	
		Promedio	DT	Promedio	DT
Introducción a la asignatura	24	14,15	2,36	5,92	2,37
Descripción del sistema de evaluación	21	13,64	2,37	6,73	3,24
Gestión del tiempo	8	9,32	1,87	6,54	2,43
Disponibilidad general	9	9,50	1,70	5,85	3,04
Coherencia organizativa	19	13,34	1,66	6,82	2,91
Implementación del sistema de evaluación	11	11,61	1,33	7,16	3,00
Resolución de dudas	14	12,03	1,83	6,17	3,11
Capacidad explicativa	5	7,45	2,23	6,01	3,21
Facilidad de seguimiento	12	9,15	1,68	4,84	2,79
Satisfacción general	12	11,23	2,49	5,41	3,00
Conjunto cuestionario	135	111,41	7,46	65,45	19,3

Nota: ^a Tiempo medido en segundos. ^b Extensión medida en unidades léxicas o palabras del enunciado. Grupo 1: Encuesta de evaluación del profesorado libre de incentivos. Grupo 2: Encuesta de evaluación del profesorado cuando se aplican incentivos

Fuente: Elaboración propia.

Asimismo, el diagrama de cajas del grupo 2 (figura 1) muestra una distribución claramente asimétrica, así como la existencia de diversos casos atípicos. Resulta especialmente significativa la existencia de múltiples casos atípicos por debajo del mínimo de 32,44 segundos ubicado en el primer cuartil de la distribución. Aquí encontramos encuestas en las que la inversión de tiempo en la lectura y cumplimentación del cuestionario en su conjunto no excedió de los veinte segundos. Valor este muy por debajo del promedio de 111,41 segundos estimado para el grupo 1 cuando no existe incentivo.

Por su parte, el análisis de pruebas paramétricas a través del estadístico t-Student indica que el tiempo de cumplimentación del cuestionario Likert difiere de forma significativa en los grupos 1 y 2. Considerando un umbral de significatividad $\alpha = 0,05$, en consecuencia, un intervalo de confianza del 95% y comprobado el supuesto de homocedasticidad con la prueba de Levene, el valor $t = 6,763$ arroja un p-valor = 0,000. El p-valor $< 0,05$ confirma la existencia de diferencias significativas en los tiempos de cumplimentación entre ambos grupos (cuadro 3).

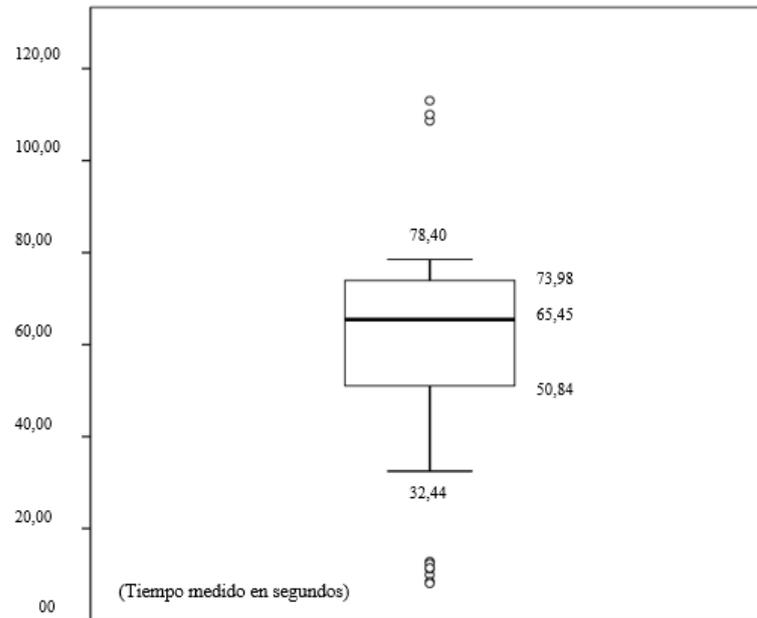


Figura 1. Distribución de tiempos invertidos en la cumplimentación de un cuestionario Likert en el grupo 2
Fuente: Elaboración propia.

Cuadro 3. Significatividad de las diferencias de tiempo empleado en un cuestionario Likert en ambos grupos

	PRUEBA DE LEVENE		PRUEBA T-STUDENT PARA LA IGUALDAD DE MEDIAS				
	F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	Error típ. de la diferencia
Se han asumido varianzas iguales	11,754	0,003	6,763	18	0,000	4,99700	0,73887
No se han asumido varianzas iguales			6,763	10,747	0,000	4,99700	0,73887

Fuente: Elaboración propia.

Los resultados de esta prueba refuerzan los datos obtenidos en el análisis descriptivo, donde observamos que el tiempo promedio de cumplimentación en el grupo en el que no se aplica incentivo por participación ($M = 111,41$, $DT = 7,46$) es manifiestamente superior al reflejado en el grupo en el que la realización de la encuesta esta incentivada ($M = 65,45$, $DT = 19,36$).

Por último, el análisis de correlaciones entre número de palabras empleadas en la formulación del enunciado en la escala Likert y el promedio de tiempo empleado en la lectura y contestación de la pregunta indica fuertes correlaciones en el primer grupo ($r = ,93$, $p < ,01$), pero correlaciones débiles y no significativas en el segundo (cuadro 4).

Esta ausencia de correlación entre las variables extensión del enunciado y promedio de tiempo, en el grupo en el que se aplica el incentivo por cumplimentación, queda patente también en los datos expuestos en el cuadro 2. En esta, observamos cómo el ítem

introducción a la asignatura, el de mayor extensión (veinticuatro palabras), presenta un tiempo promedio de cumplimentación de 5,92 segundos. Por el contrario, el ítem de menor longitud (cinco palabras), capacidad explicativa, refleja un promedio de tiempo invertido superior al anterior con 6,01 segundos.

Cuadro 4. Correlaciones entre extensión del enunciado e inversión de tiempo en un cuestionario Likert para ambos grupos

		GRUPO 1 ^a	GRUPO 2 ^a
		Promedio	Promedio
Extensión enunciado ^b	Correlación de Pearson (<i>r</i>)	0,93**	0,17
	Sig. (bilateral)	0,00	0,63
	<i>N</i>	10	10

** La correlación es significativa al nivel ,01 (bilateral). ^a Tiempo medido en segundos. ^b Extensión medida en unidades léxicas o palabras del enunciado.

Fuente: Elaboración propia.

3.2. Inversión de tiempo experimentada con un cuestionario con escalas BARS incentivado

Los resultados revelan cómo el tiempo promedio invertido por el estudiante al cumplimentar el cuestionario BARS en un escenario de incentivo (325,62 segundos) es ligeramente superior al tiempo promedio en una situación libre de recompensas (317,85 segundos). Asimismo, el promedio de tiempo invertido por los estudiantes del grupo 2 en cada pregunta se sitúa por encima del estimado para el grupo 1 en cinco ítems. Esto es así en los ítems: introducción a la asignatura, disponibilidad general, capacidad explicativa, facilidad de seguimiento y satisfacción general. En los cinco ítems restantes el promedio computado en el grupo 2 es solo levemente inferior al presentado en el grupo 1 (cuadro 5).

Cuadro 5. Inversión de tiempo empleado en un cuestionario BARS en ambos grupos

Ítem	Extensión enunciado ^b	GRUPO 1 ^a		GRUPO 2 ^a	
		Promedio	DT	Promedio	DT
Introducción a la asignatura	282	37,77	5,89	39,67	10,67
Descripción del sistema de evaluación	233	36,95	5,01	35,84	12,26
Gestión del tiempo	157	29,83	4,18	25,83	6,44
Disponibilidad general	314	40,05	10,91	45,49	4,25
Coherencia organizativa	216	31,99	6,15	30,77	9,44
Implementación del sistema de evaluación	204	26,32	4,02	26,22	7,75
Resolución de dudas	219	32,88	5,36	31,15	7,98
Capacidad explicativa	158	25,48	2,48	26,00	7,48
Facilidad de seguimiento	230	33,56	6,64	35,12	5,28
Satisfacción general	208	27,18	1,24	29,41	7,28
Conjunto Cuestionario	2222	317,85	16,97	325,62	27,78

Nota: ^a Tiempo medido en segundos. ^b Extensión medida en unidades léxicas o palabras del enunciado. Grupo 1: Encuesta de evaluación del profesorado libre de incentivos. Grupo 2: Encuesta de evaluación del profesorado cuando se aplican incentivos

Fuente: Elaboración propia.

Por otro lado, en el diagrama de cajas del grupo 2 (figura 2) observamos una distribución prácticamente simétrica. Cabe aquí también destacar la ausencia de casos atípicos en la distribución, ni por debajo del mínimo del primer cuartil, ni por encima del valor máximo del último.

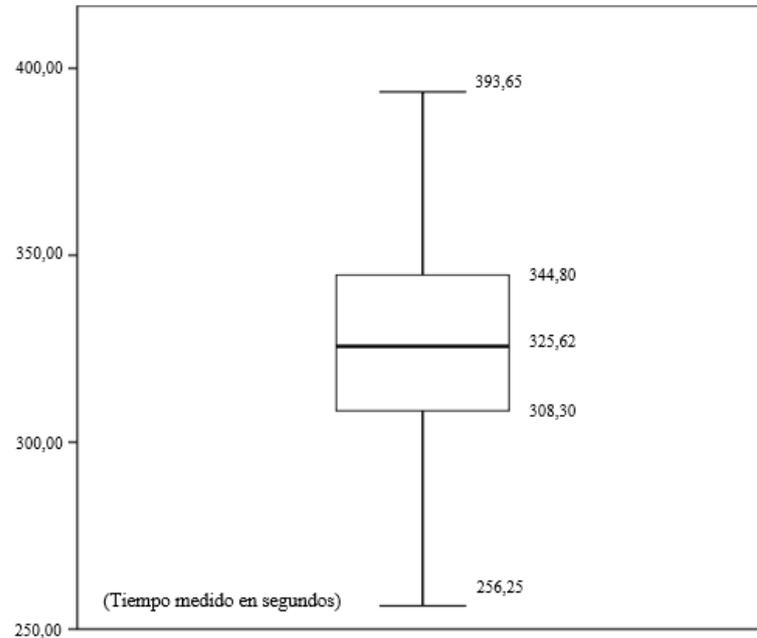


Figura 2. Distribución de tiempos invertidos en la cumplimentación de un cuestionario BARS en el grupo 2
Fuente: Elaboración propia.

Por su parte, el análisis paramétrico mediante la prueba t-Student apunta que el tiempo de realización del cuestionario BARS no difiere de manera significativa en los grupos 1 y 2. Para el umbral de significatividad fijado de $\alpha = 0,05$, un intervalo de confianza del 95% y examinada la homocedasticidad mediante la prueba de Levene, el valor $t = -0,134$ arroja un p-valor = 0,895. El p-valor > 0,05 corrobora la ausencia de diferencias de tipo significativo en los tiempos de cumplimentación entre los dos grupos de interés (cuadro 6).

Cuadro 6. Significatividad de las diferencias de tiempo empleado en un cuestionario BARS en ambos grupos

	PRUEBA DE LEVENE		PRUEBA T-STUDENT PARA LA IGUALDAD DE MEDIAS				
	F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	Error típ. de la diferencia
Se han asumido varianzas iguales	0,673	0,423	-0,134	18	0,895	-0,34900	2,59658
No se han asumido varianzas iguales			-0,134	16,956	0,895	-0,34900	2,59658

Fuente: Elaboración propia.

Este hecho puede observarse también en los datos derivados del análisis descriptivo, donde se muestra como el tiempo promedio de cumplimentación en el grupo no

incentivado ($M = 317,85$, $DT = 16,97$) es incluso menor al del grupo en el que se incentiva la cumplimentación de la encuesta ($M = 325,62$, $DT = 27,78$).

Por último, el estudio de las correlaciones entre el volumen de palabras utilizadas para la redacción de los episodios de comportamiento en la escala BARS y el promedio de tiempo invertido en leer y responder la pregunta revela correlaciones significativas en ambos grupos. La correlación entre extensión del enunciado y tiempo de cumplimentación en el primer grupo arroja un coeficiente $r = 0,86$, $p < 0,01$. De igual forma, ambas variables presentan también una correlación fuerte y significativa de $r = 0,96$, $p < 0,01$ en el segundo grupo objeto de estudio (cuadro 7).

Cuadro 7. Correlaciones entre extensión del enunciado e inversión de tiempo en un cuestionario BARS para ambos grupos

		GRUPO 1 ^a	GRUPO 2 ^a
		Promedio	Promedio
Extensión enunciado ^b	Correlación de Pearson (r)	0,86**	0,96**
	Sig. (bilateral)	0,00	0,00
	N	10	10

** La correlación es significativa al nivel ,01 (bilateral). ^a Tiempo medido en segundos. ^b Extensión medida en unidades léxicas o palabras del enunciado.

Fuente: Elaboración propia.

Las correlaciones entre ambas variables pueden constatar en la congruencia de los datos reflejados en el cuadro 5. En esta, observamos cómo la pregunta del ítem de mayor extensión (314 palabras), disponibilidad general, presenta un tiempo promedio de cumplimentación de 40,05 segundos en el grupo 1 y 45,49 segundos en el grupo 2. Por otro lado, el ítem de menor longitud (157 palabras), gestión del tiempo, refleja un promedio de 29,83 segundos en el primer grupo y 25,83 en el segundo. Por último, un ítem de longitud intermedia (219 palabras), como el ítem resolución de dudas, presenta tiempos promedio de ejecución similares en ambos grupos, 32,88 segundos en el primero y 31,15 en el segundo.

4. Conclusiones

Aunque la ausencia de una base sólida de trabajos que exploren específicamente el tema de la lectura de las encuestas de evaluación del profesorado impide el desarrollo de una discusión comparativa con otros autores, los hallazgos de la presente investigación permiten sacar varias conclusiones.

En primer lugar, los resultados revelan importantes diferencias en el tiempo invertido por el estudiante universitario para abordar las evaluaciones del profesorado, cuando se aplican incentivos por participación, en función de la tipología del cuestionario empleada.

Los hallazgos en los análisis descriptivos muestran cómo el tiempo promedio invertido en el cuestionario Likert, se ve minorado cuando se aplica un incentivo por cumplimentación en la encuesta. Por el contrario, cuando se trata de cuestionarios BARS el promedio de tiempo empleado no se reduce a pesar de la introducción de la recompensa por participación. Además, mientras la información recabada en el segundo grupo utilizando el cuestionario BARS muestra una distribución homogénea de los tiempos de cumplimentación, la información recopilada a través del cuestionario Likert presenta

múltiples valores atípicos. Estos casos atípicos, muy por debajo de los mínimos de la distribución, apuntan una falta de lectura íntegra de los enunciados propuestos en la encuesta cuando se aplican incentivos durante la evaluación.

En la misma línea, los análisis paramétricos realizados mediante la prueba t-Student, muestran que mientras en el cuestionario Likert existen diferencias significativas en los tiempos de cumplimentación en función de si se aplican o no incentivos por participación, en el cuestionario BARS el tiempo de cumplimentación no experimenta variaciones estadísticamente relevantes. Asimismo, el análisis de correlaciones bivariadas empleando el coeficiente de correlación de Pearson muestra también diferencias entre ambos cuestionarios para los dos grupos de interés. Los hallazgos realizados en el grupo 1, bajo un sistema libre de incentivos, revelan fuertes correlaciones entre la extensión del enunciado y el tiempo de cumplimentación con independencia de la tipología de escala empleada. Por el contrario, cuando el estudiante realiza la evaluación ante un escenario incentivado (Grupo 2), las correlaciones entre extensión del enunciado e inversión de tiempo difieren sustancialmente en función del tipo de cuestionario. Mientras que en los colectivos de alumnos que emplean el cuestionario BARS se mantienen las correlaciones entre ambas variables, en el colectivo en que se utiliza el cuestionario Likert no se detectan correlaciones significativas entre extensión de enunciado y tiempo invertido.

Los autores, partiendo de una premisa de condiciones experimentales constantes en las muestras analizadas, consideran que los datos de tiempo de cumplimentación y las correlaciones entre extensión del enunciado e inversión de tiempo, unidos a la presencia de casos atípicos en el grupo 2, apuntan hacia la existencia de una falta de atención del alumno en sistemas de evaluación incentivados en los que se emplean cuestionarios Likert. Bajo la perspectiva de los investigadores, las diferencias observadas en los cuestionarios Likert, pueden ocasionarse por diferentes motivos y tener distintas implicaciones. En lo que atañe a los motivos, los autores, en línea con lo apuntado por estudios previos (Cañadas y Cuétara, 2018; Moreno Olivos, 2018), señalan que las evaluaciones con estos cuestionarios pueden ocasionar la aparición de un componente de cansancio derivado del carácter repetitivo de la escala. Asimismo, los investigadores, en consonancia con lo apuntado por Franklin (2001) o McClain, Gulbis y Hays (2018), alertan del riesgo de inflamiento de las puntuaciones del docente en este tipo de cuestionario.

En opinión de los autores, cuando de escalas Likert se trata, tanto el factor cansancio como el inflamiento de las puntuaciones pueden guardar vinculación con los reducidos tiempos de cumplimentación de la encuesta cuando se introducen incentivos. Los resultados revelan, por tanto, que la aplicación de incentivos en la realización de encuestas de evaluación del profesorado, aun posibilitando mejoras en las tasas de participación, también puede tener implicaciones negativas para determinadas tipologías de cuestionario.

En opinión de los autores, la introducción de recompensas puede convertir una tarea de elevada exigencia en términos de atención en un proceso mecánico que el estudiante aborda con celeridad en su afán por alcanzar el beneficio prometido. Esto lleva al alumno a proporcionar respuestas aleatorias, e incluso contestaciones sin llegar a leer íntegramente los enunciados, perjudicando así la significatividad de la información recogida. Esta situación desvirtúa los resultados obtenidos y, en línea con lo apuntado por Moreno Olivos (2018), distorsiona el sentido original de la evaluación.

La utilización de cuestionarios BARS lleva al estudiante a realizar una lectura más detallada de los enunciados de cada ítem. Este hecho exige necesariamente aumentar el nivel de atención, lo cual permite reducir las implicaciones negativas de la aplicación de incentivos durante la evaluación. Los cuestionarios BARS, a diferencia de los cuestionarios con escalas de Likert, implican activamente a los potenciales futuros evaluadores en el diseño de la escala (Dickinson y Zellinger, 1980; Fernández Millán y Fernández Navas, 2013). Tanto es así que son los propios estudiantes los que, durante el proceso de construcción, proporcionan, normalmente a través de entrevistas, los episodios de comportamiento que a posteriori se utilizarán para conformar el instrumento de medición final (Matosas-López, Leguey-Galán y Leguey-Galán, 2019). Los autores, en línea con Martin-Raugh y otros (2016), consideran que el uso de episodios de comportamiento en los cuestionarios BARS favorece la implicación del estudiante durante la cumplimentación de la encuesta; dado que este reconoce en el cuestionario escenarios familiares que le ayudan a retrotraerse a situaciones reales observadas en la labor docente a lo largo del curso. Otro de los rasgos identificativos de los cuestionarios BARS que, en opinión de los autores, redundan positivamente en una adecuada cumplimentación de la encuesta es la utilización del vocabulario y terminología del evaluador en la redacción de los enunciados usados en la escala (Harari y Zedeck, 1973; Jacobs, Kafry y Zedeck, 1980). Hecho este que facilita la identificación y acercamiento del estudiante al proceso de evaluación.

Todo esto, puede llegar a favorecer que la inversión de tiempo para completar el cuestionario BARS no se vea alterada, aun cuando el alumno aborda la evaluación en un escenario sujeto a incentivos. Por lo tanto, los autores concluyen que, aunque los instrumentos de evaluación del profesorado con escalas de Likert no parecen favorecer una óptima lectura y cumplimentación de la encuesta cuando la evaluación introduce recompensas, esta situación puede ser mejorada empleando cuestionarios BARS.

Considerando el peso de las encuestas de evaluación del profesorado; no solo en lo que atañe a la mejora de la calidad docente (fin formativo), sino también en lo que a la promoción del profesorado se refiere (fin sumativo) (Linse, 2017; Nygaard y Belluigi, 2011); la adecuada lectura de los cuestionarios es un asunto que no puede ser obviado por la comunidad académica.

Limitaciones e investigaciones futuras

El presente trabajo, sin embargo, no está exento de limitaciones. La contextualización de la investigación en una única universidad y facultad invita a reflexionar sobre si estudiantes de otras instituciones o de disciplinas ajenas al campo de las ciencias sociales pueden mostrar mejor predisposición hacia los procesos de evaluación del profesorado cuando se aplican incentivos por participación. Asimismo, el estudio queda circunscrito a la etapa de grado, con lo cual no podemos inferir que estudiantes de máster y posgrado - estudios sometidos también a este tipo de evaluaciones- vayan a actuar de idéntica manera. Ambos aspectos invitan a tomar los resultados con cautela y recomiendan huir de generalizaciones sobre el conjunto del sistema universitario.

Sin embargo, a pesar de las limitaciones señaladas, el presente estudio no solo arroja luz sobre un problema prácticamente ignorado por investigaciones previas, sino que, además, aborda el mismo proponiendo alternativas. Asimismo, este trabajo abre nuevas vías de estudio en el campo, postulando los cuestionarios con episodios de comportamiento BARS como opción a considerar en los procesos de evaluación del profesorado.

Referencias

- Abrami, P. C. y D'Apollonia, S. (1997). Navigating student ratings of instruction. *American Psychologist*, 52(11), 1198-1208. Recuperado de <http://psycnet.apa.org/buy/1997-43129-004>
- Ballantyne, C. (2003). Online evaluations of teaching: An examination of current practice and considerations for the future. *New Directions for Teaching and Learning*, 96, 103-112. <https://doi.org/10.1002/tl.127>
- Bernardin, H. J. (1977). Behavioural expectation scales versus summated scales. *Journal of Applied Psychology*, 62(4), 422-427. Recuperado de <http://psycnet.apa.org/record/1978-09104-001>
- Boring, A. (2017). Gender biases in student evaluations of teaching. *Journal of Public Economics*, 145, 27-41. <https://doi.org/10.1016/j.jpubeco.2016.11.006>
- Buendía, L. (1997). La investigación por encuesta. En L. Buendía, P. Colás y F. Hernández Pina (Eds.), *Métodos de investigación en psicopedagogía* (pp. 120-154). Madrid: McGraw-Hill.
- Buendía, L. (1994). El proceso de investigación. En L. Buendía y P. Colás (Eds.), *Investigación educativa* (pp. 69-108). Sevilla: Alfar.
- 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>
- Darwin, S. (2017). What contemporary work are student ratings actually doing in higher education? *Studies in Educational Evaluation*, 54, 13-21. <https://doi.org/10.1016/j.stueduc.2016.08.002>
- De-Juanas Oliva, A. y Beltrán Llera, J. A. (2013). Valoraciones de los estudiantes de ciencias de la educación sobre la calidad de la docencia universitaria. *Educación XX1*, 17(1), 59-82. <https://doi.org/10.5944/educxx1.17.1.10705>
- Dickinson, T. L. y Zellinger, P. M. (1980). A comparison of the behaviorally anchored rating and mixed standard scale formats. *Journal of Applied Psychology*, 65(2), 147-154. <https://doi.org/10.1037//0021-9010.65.2.147>
- Dommeier, 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>
- Escobar-Pérez, J. y Cuervo-Martínez, Á. (2008). Validez de contenido y juicio de expertos: Una aproximación a su utilización. *Avances en Medición*, 6, 27-36.
- Feistauer, D. y Richter, T. (2016). How reliable are students' evaluations of teaching quality? A variance components approach. *Assessment y Evaluation in Higher Education*, 47(8), 1-17. <https://doi.org/10.1080/02602938.2016.1261083>
- Feldman, K. A. (1978). Course characteristics and college students' ratings of their teachers: What we know and what we don't. *Research in Higher Education*, 9(3), 199-242. <https://doi.org/10.1007/BF00976997>
- Fernández Millán, J. M. y Fernández Navas, M. (2013). Elaboración de una escala de evaluación de desempeño para educadores sociales en centros de protección de menores. *Intangible Capital*, 9(3), 571-589. <https://doi.org/10.3926/ic.410>

- Ficapal-Cusí, P., Torrent-Sellens, J., Boada-Grau, J. y Sánchez-García, J.-C. (2013). Evaluación del e-learning en la formación para el empleo: Estructura factorial y fiabilidad. *Revista de Educación*, 361, 9-7. <https://doi.org/10.4438/1988-592X-RE-2013-361-232>
- Franklin, J. (2001). Interpreting the numbers: Using a narrative to help others read student evaluations of your teaching accurately. *New Directions for Teaching and Learning*, 87, 85-100. <https://doi.org/10.1002/tl.10001>
- Galbraith, C. S. y Merrill, G. B. (2012). Predicting student achievement in university-level business and economics classes: Peer observation of classroom instruction and student ratings of teaching effectiveness. *College Teaching*, 60(2), 48-55. <https://doi.org/10.1080/87567555.2011.627896>
- Gannaway, D., Green, T. y Mertova, P. (2017). So how big is big? Investigating the impact of class size on ratings in student evaluation. *Assessment y Evaluation in Higher Education*, 8(2), 1-10. <https://doi.org/10.1080/02602938.2017.1317327>
- George, D. y Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference*. Los Ángeles, CA: Allyn and Bacon.
- Griffin, B. W. (2004). Grading leniency, grade discrepancy, and student ratings of instruction. *Contemporary Educational Psychology*, 29(4), 410-425. <https://doi.org/10.1016/J.CEDPSYCH.2003.11.001>
- Guzmán, J. C. (2018). Las buenas prácticas de enseñanza de los profesores de educación superior. *REICE. Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 16(2), 133-149. <https://doi.org/10.15366/reice2018.16.2.008>
- Harari, O. y Zedeck, S. (1973). Development of behaviorally anchored scales for the evaluation of faculty teaching. *Journal of Applied Psychology*, 58(2), 261-265. <https://doi.org/10.1037/h0035633>
- Hernández Pina, F. (1997). Diseños de investigación experimental. En L. Buendía, P. Colás y F. Hernández Pina (Eds.), *Métodos de investigación en psicopedagogía* (pp. 91-117). Madrid: McGraw-Hill.
- Jacobs, R., Kafry, D. y Zedeck, S. (1980). Expectations of behaviorally anchored rating scales. *Personnel Psychology*, 33(3), 595-640. <https://doi.org/10.1111/j.1744-6570.1980.tb00486.x>
- Johnson, T. D. (2003). Online student ratings: Will students respond? *New Directions for Teaching and Learning*, 96, 49-59. <https://doi.org/doi:10.1002/tl.122>
- Linse, A. R. (2017). Interpreting and using student ratings data: Guidance for faculty serving as administrators and on evaluation committees. *Studies in Educational Evaluation*, 54, 94-106. <https://doi.org/10.1016/j.stueduc.2016.12.004>
- Lizasoain, L., Etxebarria, J. y Lukas, J. F. (2017). Propuesta de un nuevo cuestionario de evaluación de los profesores de la Universidad del País Vasco. Estudio psicométrico, dimensional y diferencial. *RELIEVE. Revista Electrónica de Investigación y Evaluación Educativa*, 23(1), 1-21. <https://doi.org/10.7203/relieve.23.2.10436>
- Luna Serrano, E. (2015). Validación de constructo de un cuestionario de evaluación de la competencia docente. *Revista Electronica de Investigación Educativa*, 17(3), 27-45.
- Marsh, W. (1982). SEEQ: A reliable, valid, and useful instrument for collecting students' evaluations of university teaching. *British Journal of Educational Psychology*, 52(2), 77-95. <https://doi.org/10.1111/j.2044-8279.1982.tb02505.x>

- Marsh, W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research*, 11(3), 253-388. [https://doi.org/10.1016/0883-0355\(87\)90001-2](https://doi.org/10.1016/0883-0355(87)90001-2)
- Marsh, W. (1991). A multidimensional perspective on students' evaluations of teaching effectiveness-reply to Abrami and Dapollonia (1991). *Journal of Educational Psychology*, 83(3), 416-421. <https://doi.org/10.1037//0022-0663.83.3.416>
- Martin-Raugh, M., Tannenbaum, R. J., Tocci, C. M. y Reese, C. (2016). Behaviourally 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>
- Matosas-López, L. y Leguey-Galán, S. (2018). Implementación de behavioral anchored rating scales (BARS) para la evaluación del profesorado universitario en asignaturas de modalidad online. En C. Monge López, P. Gómez Hernández y R. Herrero Marcos (Eds.), *Actas del I Congreso Virtual Internacional y III Congreso Virtual Iberoamericano sobre Recursos Educativos Innovadores CIREI* (pp. 204-208). Madrid: Fundación General de la Universidad de Alcalá.
- Matosas-López, L., Aguado-Franco, J. C. y Gómez-Galán, J. (2019). Constructing an instrument with behavioral scales to assess teaching quality in blended learning modalities. *Journal of New Approaches in Educational Research*, 8(2).
- Matosas-López, L., Leguey-Galán, S. y Leguey-Galán, S. (2019). Evaluación de la calidad y la eficiencia docente en el contexto de la educación superior: Alternativas de mejora. En J. Gómez-Galán, A. Martín-Padilla y H. Cobos (Ed.), *La educación superior en el siglo XXI: Una mirada multidisciplinaria* (pp. 240-257). Wheaton, IL: Editorial UMET.
- Mayorga Fernández, M. J. y Ruiz Baeza, V. M. (2002). Muestreos utilizados en investigación educativa en España. *RELIEVE. Revista Electrónica de Investigación y Evaluación Educativa*, 8(2), 195-165.
- McCann, S. y Gardner, C. (2014). Student personality differences are related to their responses on instructor evaluation forms. *Assessment y Evaluation in Higher Education*, 39(4), 1-15. <https://doi.org/10.1080/02602938.2013.845647>
- McClain, L., Gulbis, A. y Hays, D. (2018). Honesty on student evaluations of teaching: Effectiveness, purpose, and timing matter! *Assessment and Evaluation in Higher Education*, 43(3), 369-385. <https://doi.org/10.1080/02602938.2017.1350828>
- McPherson, M. A. (2006). Determinants of how students evaluate teachers. *The Journal of Economic Education*, 37(1), 3-20. <https://doi.org/10.3200/JECE.37.1.3-20>
- Molero López-Barajas, D. M. y Ruiz Carrascosa, J. (2005). La evaluación de la docencia universitaria. Dimensiones y variables más relevantes. *Revista de Investigación Educativa*, 23(1), 57-84. Recuperado de <http://revistas.um.es/rie/article/view/98341>
- 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>
- Morley, D. D. (2012). Claims about the reliability of student evaluations of instruction: The ecological fallacy rides again. *Studies in Educational Evaluation*, 38(1), 15-20. <https://doi.org/10.1016/j.stueduc.2012.01.001>
- Muñoz Cantero, J. M., Ríos De Deus, M. P. y Abalde Paz, E. (2002). Evaluación docente vs evaluación de la calidad. *RELIEVE*, 8(2), 103-134.
- Nair, C. S. y Adams, P. (2009). Survey platform: A factor influencing online survey delivery and response rate. *Quality in Higher Education*, 15(3), 291-296. <https://doi.org/10.1080/13538320903399091>

- Nasser-Abu Alhija, F. y Fresko, B. (2009). Student evaluation of instruction: What can be learned from students' written comments? *Studies in Educational Evaluation*, 35(1), 37-44.
<https://doi.org/10.1016/j.stueduc.2009.01.002>
- 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(3), 301-314.
<https://doi.org/10.1080/02602930701293231>
- Nygaard, C. y Belluigi, D. Z. (2011). A proposed methodology for contextualised evaluation in higher education. *Assessment y Evaluation in Higher Education*, 36(6), 657-671.
<https://doi.org/10.1080/02602931003650037>
- Reyero, D. (2014). La excelencia docente universitaria. Análisis y propuestas para una mejor evaluación del profesorado universitario. *Educación XXI*, 17(2), 125-143.
<https://doi.org/10.5944/educxx1.17.2.11482>
- Ruiz Carrascosa, J. (2000). La evaluación de la enseñanza por los alumnos en el plan nacional de evaluación de la calidad de las universidades. Construcción de un instrumento de valoración. *Revista de Investigación Educativa*, 18(2), 433-445.
- Sharon, A. T. y Bartlett, C. J. (1969). Effect of instructional conditions in producing leniency on two types of rating scales. *Personnel Psychology*, 22(3), 251-263.
<https://doi.org/10.1111/j.1744-6570.1969.tb00330.x>
- Sorenson, D. L. y Reiner, C. (2003). Charting the uncharted seas of online student ratings of instruction. *New Directions for Teaching and Learning*, 96(1), 1-24.
<https://doi.org/10.1002/tl.118>
- Spooren, P. (2010). On the credibility of the judge. A cross-classified multilevel analysis on students' evaluation of teaching. *Studies in Educational Evaluation*, 36(4), 121-131.
<https://doi.org/10.1016/j.stueduc.2011.02.001>
- Spooren, P., Mortelmans, D. y Christiaens, W. (2014). Assessing the validity and reliability of a quick scan for student's evaluation of teaching. Results from confirmatory factor analysis and G theory. *Studies in Educational Evaluation*, 43, 88-94.
<https://doi.org/10.1016/j.stueduc.2014.03.001>
- Stanny, C. J. y Arruda, J. E. (2017). A comparison of student evaluations of teaching with online and paper-based administration. *Scholarship of Teaching and Learning in Psychology*, 3(3), 198-207. <https://doi.org/10.1037/stl0000087>
- Stoskopf, C. H., Glik, D. C., Baker, S. L., Ciesla, J. R. y Cover, C. M. (1992). The reliability and construct validity of a behaviorally anchored rating scale used to measure nursing assistant performance. *Evaluational Review*, 16(3), 333-345.
- 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>
- Tejedor Tejedor, F. J. (2009). Evaluación del profesorado universitario: Enfoque metodológico y algunas aportaciones de la investigación. *Estudios sobre Educación*, 16, 79-102.

Anexo I. Cuestionario Likert

A continuación, valorarás el trabajo desarrollado por tu profesor/a en diez categorías de la labor docente, utilizando una escala de 1 a 5. Siendo: 1=Muy en desacuerdo, 2=En desacuerdo, 3=Ni de acuerdo ni en desacuerdo, 4=De acuerdo, 5=Muy de acuerdo.

1.- El profesor/a explica detalladamente en sus primeros días de clase la guía docente de la asignatura, o la parte del temario que imparte.

1 2 3 4 5

2.- El profesor/a ha informado claramente sobre los criterios de evaluación de la asignatura (o de la parte del temario que imparte).

1 2 3 4 5

3.- El profesor/a respeta los horarios de clase.

1 2 3 4 5

4.- El profesor/a está disponible para atender a los/-as alumnos/-as.

1 2 3 4 5

5.- Las actividades docentes se ajustan a los objetivos, contenidos y metodología especificados en la guía docente de la asignatura.

1 2 3 4 5

6.- La evaluación se ajusta a las especificaciones de la guía docente.

1 2 3 4 5

7.- El profesor/a aclara adecuadamente las dudas de las distintas actividades propuestas en la asignatura.

1 2 3 4 5

8.- El profesor/a explica con claridad.

1 2 3 4 5

9.- El desarrollo de las clases me permite un seguimiento y aprendizaje adecuados.

1 2 3 4 5

10.- Estoy satisfecho/-a con la labor que desarrolla el profesor o la profesora.

1 2 3 4 5

Anexo II. Cuestionario BARS

Señala el conjunto de episodios de comportamiento que mejor define el desempeño de tu profesor/a para cada una de las diez siguientes categorías de la labor docente.

1.- Introducción a la asignatura

- El profesor/a NO presenta todos los puntos clave de la guía docente al inicio del curso (programa de estudios, competencias, objetivos, metodología de trabajo, ECTS, recursos en Aula Virtual ...); NO expone la importancia de la asignatura en términos académicos / profesionales; NO describe el plan cronológico de contenidos de la asignatura ni la inversión de tiempo requerida en cada parte (clases, exámenes, tareas ...) y NO proporciona una descripción detallada de la bibliografía / materiales de apoyo, ni de la forma de usarlos
- El profesor/a expone la importancia de la asignatura en términos académicos / profesionales
- El profesor/a presenta todos los puntos clave de la guía docente al inicio del curso (programa de estudios, competencias, objetivos, metodología de trabajo, ECTS, recursos en Aula Virtual ...) y proporciona una descripción detallada de la bibliografía / materiales de apoyo, además de la forma de usarlos
- El profesor/a presenta todos los puntos clave de la guía docente al inicio del curso (programa de estudios, competencias, objetivos, metodología de trabajo, ECTS, recursos en Aula Virtual ...); expone la importancia de la asignatura en términos académicos / profesionales y proporciona una descripción detallada de la bibliografía / materiales de apoyo, además de la forma de usarlos
- El profesor/a presenta todos los puntos clave de la guía docente al inicio del curso (programa de estudios, competencias, objetivos, metodología de trabajo, ECTS, recursos en Aula Virtual ...); expone la importancia de la asignatura en términos académicos / profesionales; describe el plan cronológico de contenidos de la asignatura y la inversión de tiempo requerida en cada parte (clases, exámenes, tareas ...) y proporciona una descripción detallada de la bibliografía / materiales de apoyo, además de la forma de usarlos

2.- Descripción del sistema de evaluación

- El profesor/a NO explica todos los puntos clave del sistema de evaluación (nº de exámenes, fechas de exámenes, % de evaluación teórica / práctica, tareas, cuestionarios, exámenes de recuperación ...); NO expone los aspectos específicos relacionados con los potenciales exámenes o pruebas de evaluación (material de apoyo, criterios de corrección, duración del examen ...); NO enumera los principales contenidos sujetos a evaluación y NO explica la forma en que se consideran la participación en clase, la asistencia o las actividades complementarias en la evaluación de la asignatura
- El profesor/a enumera los principales contenidos sujetos a evaluación y explica la forma en que se consideran la participación en clase, la asistencia o las actividades complementarias en la evaluación de la asignatura
- El profesor/a describe los principales contenidos sujetos a evaluación; explica la forma en que se consideran la participación en clase, la asistencia o las actividades complementarias en la evaluación de la asignatura y expone los

aspectos específicos relacionados con los potenciales exámenes o pruebas de evaluación (material de apoyo, criterios de corrección, duración del examen ...)

- El profesor/a explica todos los puntos clave del sistema de evaluación (nº de exámenes, fechas de exámenes, % de evaluación teórica / práctica, tareas, cuestionarios, exámenes de recuperación ...); expone los aspectos específicos relacionados con los potenciales exámenes o pruebas de evaluación (material de apoyo, criterios de corrección, duración del examen ...) y enumera los principales contenidos sujetos a evaluación
- El profesor/a explica todos los puntos clave del sistema de evaluación (nº de exámenes, fechas de exámenes, % de evaluación teórica / práctica, tareas, cuestionarios, exámenes de recuperación ...); expone los aspectos específicos relacionados con los potenciales exámenes o pruebas de evaluación (material de apoyo, criterios de corrección, duración del examen ...); enumera los principales contenidos sujetos a evaluación y explica la forma en que se consideran la participación en clase, la asistencia o las actividades complementarias en la evaluación de la asignatura

3.- Gestión del tiempo

- El profesor/a notifica por adelantado cualquier cambio en los horarios de clase o las potenciales ausencias cuando es necesario
- El profesor/a mantiene un tiempo de docencia homogéneo diariamente / semanalmente
- El profesor/a notifica por adelantado cualquier cambio en los horarios de clase o las potenciales ausencias cuando es necesario y mantiene un tiempo de docencia homogéneo diariamente / semanalmente
- El profesor/a notifica por adelantado cualquier cambio en los horarios de clase o las potenciales ausencias cuando es necesario y es puntual en la llegada a clase para preparar los materiales didácticos requeridos (notas, proyecciones, recursos multimedia ...)
- El profesor/a es puntual en la llegada a clase para preparar los materiales didácticos requeridos (notas, proyecciones, recursos multimedia ...); administra eficazmente el tiempo de clase, recuperando retrasos o posibles ausencias si es necesario; notifica por adelantado cualquier cambio en los horarios de clase o las potenciales ausencias cuando es necesario y mantiene un tiempo de docencia homogéneo diariamente / semanalmente

4.- Disponibilidad general

- El profesor/a NO define un horario de tutorías adecuado para los estudiantes (turno de mañana / tarde, la misma ubicación donde se imparte la asignatura ...); NO está abierto para ayudar o atender a los estudiantes antes / después de las clases; NO presenta diferentes canales de contacto ni la forma o los momentos de uso de cada uno de ellos (eMail, Aula Virtual, teléfono, despacho del profesor/a...) y NO responde a las consultas en remoto de los estudiantes (eMail, Aula Virtual, Videoconferencia) en un periodo de 48-72 horas
- El profesor/a presenta diferentes canales de contacto además de la forma y los momentos de uso de cada uno de ellos (eMail, Aula Virtual, teléfono, despacho del profesor/a...)

- El profesor/a presenta diferentes canales de contacto además de la forma y los momentos de uso de cada uno de ellos (eMail, Aula Virtual, teléfono, despacho del profesor/a...) y responde a las consultas en remoto de los estudiantes (eMail, Aula Virtual, Videoconferencia) en un máximo de 48-72 horas
- El profesor/a está abierto para ayudar o atender a los estudiantes antes / después de las clases; presenta diferentes canales de contacto además de la forma y los momentos de uso de cada uno de ellos (eMail, Aula Virtual, teléfono, despacho del profesor/a...) y responde a las consultas en remoto de los estudiantes (eMail, Aula Virtual, Videoconferencia) en un máximo de 48-72 horas
- El profesor/a define un horario de tutorías adecuado para los estudiantes (turno de mañana / tarde, la misma ubicación donde se imparte la asignatura ...); está abierto para ayudar o atender a los estudiantes antes / después de las clases; presenta diferentes canales de contacto además de la forma y los momentos de uso de cada uno de ellos (eMail, Aula Virtual, teléfono, despacho del profesor/a...) y responde a las consultas en remoto de los estudiantes (eMail, Aula Virtual, Videoconferencia) en un máximo de 48-72 horas.

5.- Coherencia organizativa

- El profesor/a NO mantiene la metodología de trabajo inicialmente presentada al inicio del curso; NO respeta el plan cronológico de contenidos diseñado para cada semana; NO desarrolla exactamente el contenido en el programa de la asignatura (Ni más ni menos contenido) y NO prioriza los contenidos clave para alcanzar los objetivos / competencias de la asignatura ni indica las competencias específicas desarrolladas en cada actividad formativa
- El profesor/a respeta el plan cronológico de contenidos diseñado para cada semana
- El profesor/a mantiene la metodología de trabajo inicialmente presentada al inicio del curso y respeta el plan cronológico de contenidos diseñado para cada semana
- El profesor/a mantiene la metodología de trabajo inicialmente presentada al inicio del curso; desarrolla exactamente el contenido en el programa de la asignatura (Ni más ni menos contenido) y prioriza los contenidos clave para alcanzar los objetivos / competencias de la asignatura e indica las competencias específicas desarrolladas en cada actividad formativa
- El profesor/a mantiene la metodología de trabajo inicialmente presentada al inicio del curso; respeta el plan cronológico de contenidos diseñado para cada semana; desarrolla exactamente el contenido en el programa de la asignatura (Ni más ni menos contenido) y prioriza los contenidos clave para alcanzar los objetivos / competencias de la asignatura e indica las competencias específicas desarrolladas en cada actividad formativa

6.- Implementación del sistema de evaluación

- El profesor/a NO mantiene una coherencia con el método de evaluación / criterios de evaluación previamente descritos; NO respeta las fechas de los exámenes fijados al inicio del curso ni mantiene las fechas institucionales; NO realiza el número de exámenes originalmente previsto y NO concentra las

actividades de evaluación en torno a los materiales impartidos durante la asignatura

- El profesor/a respeta las fechas de los exámenes fijados al inicio del curso y / o mantiene las fechas institucionales
- El profesor/a respeta las fechas de los exámenes fijados al inicio del curso y / o mantiene las fechas institucionales y realiza el número de exámenes originalmente previsto
- El profesor/a mantiene una coherencia con el método de evaluación / criterios de evaluación previamente descritos; respeta las fechas de los exámenes fijados al inicio del curso y / o mantiene las fechas institucionales y realiza el número de exámenes originalmente previsto
- El profesor/a mantiene una coherencia con el método de evaluación / criterios de evaluación previamente descritos; respeta las fechas de los exámenes fijados al inicio del curso y / o mantiene las fechas institucionales; realiza el número de exámenes originalmente previsto y concentra las actividades de evaluación en torno a los materiales impartidos durante la asignatura

7.- Resolución de dudas

- El profesor/a **NO** genera un ambiente adecuado que anime a los estudiantes a formular dudas o compartir opiniones; **NO** establece pausas específicas durante las clases para formular dudas; **NO** es capaz de hacer frente a las dudas de los estudiantes presentando un solo concepto de varias maneras y **NO** resuelve las dudas de los estudiantes usando ejemplos prácticos / materiales de apoyo para afianzar la idea
- El profesor/a establece pausas específicas durante las clases para formular dudas
- El profesor/a genera un ambiente adecuado que anima a los estudiantes a formular dudas o compartir opiniones y establece pausas específicas durante las clases para formular dudas
- El profesor/a genera un ambiente adecuado que anima a los estudiantes a formular dudas o compartir opiniones; es capaz de hacer frente a las dudas de los estudiantes presentando un solo concepto de varias maneras y resuelve las dudas de los estudiantes usando ejemplos prácticos / materiales de apoyo para afianzar la idea
- El profesor/a genera un ambiente adecuado que anima a los estudiantes a formular dudas o compartir opiniones; establece pausas específicas durante las clases para formular dudas; es capaz de hacer frente a las dudas de los estudiantes presentando un solo concepto de varias maneras y resuelve las dudas de los estudiantes usando ejemplos prácticos / materiales de apoyo para afianzar la idea

8.- Capacidad explicativa

- El profesor/a utiliza recursos multimedia (diapositivas, vídeos, sitios web ...) además de la pizarra para apoyar las explicaciones
- El profesor/a utiliza recursos multimedia (diapositivas, vídeos, sitios web ...) además de la pizarra para apoyar las explicaciones y aplica un enfoque teórico-práctico que estimula el aprendizaje

- El profesor/a presenta contenidos de forma clara y concreta y utiliza recursos multimedia (diapositivas, vídeos, sitios web ...) además de la pizarra para apoyar las explicaciones
- El profesor/a presenta contenidos de forma clara y concreta; utiliza recursos multimedia (diapositivas, vídeos, sitios web ...) además de la pizarra para apoyar las explicaciones y aplica un enfoque teórico-práctico que estimula el aprendizaje
- El profesor/a presenta contenidos de forma clara y concreta; utiliza recursos multimedia (diapositivas, vídeos, sitios web ...) además de la pizarra para apoyar las explicaciones; aplica un enfoque teórico-práctico que estimula el aprendizaje y utiliza habilidades de comunicación verbal (tono, ritmo ...) / no verbal (gesto, movimiento ...) apropiadas para facilitar la comprensión

9.- Facilidad de seguimiento

- El profesor/a NO conecta los contenidos a lo largo de las etapas del curso creando una perspectiva general del tema; NO asigna una carga de trabajo semanal / mensual manejable para el estudiante; NO anima al estudiante a participar en el curso de diferentes maneras (trabajo de clase, preguntas de clase, foros de discusión on-line ...) y NO resume las principales ideas expuestas en clase diariamente / semanalmente
- El profesor/a asigna una carga de trabajo semanal / mensual manejable para el estudiante
- El profesor/a conecta los contenidos a lo largo de las etapas del curso creando una perspectiva general del tema y asigna una carga de trabajo semanal / mensual manejable para el estudiante
- El profesor/a conecta los contenidos a lo largo de las etapas del curso creando una perspectiva general del tema; anima al estudiante a participar en el curso de diferentes maneras (trabajo de clase, preguntas de clase, foros de discusión on-line ...) y resume las principales ideas expuestas en clase diariamente / semanalmente
- El profesor/a conecta los contenidos a lo largo de las etapas del curso creando una perspectiva general del tema; asigna una carga de trabajo semanal / mensual manejable para el estudiante; anima al estudiante a participar en el curso de diferentes maneras (trabajo de clase, preguntas de clase, foros de discusión on-line ...) y resume las principales ideas expuestas en clase diariamente / semanalmente

10.- Satisfacción general

- El profesor/a NO contribuye decisivamente a la consecución de los objetivos / competencias esperados en la asignatura; NO influye en el desarrollo académico / profesional del estudiante; NO demuestra un conocimiento reconocido en el campo ni la capacidad de transmitir ese conocimiento y NO es capaz de aumentar el interés de los estudiantes en ese campo de formación
- El profesor/a demuestra un conocimiento reconocido en el campo además de la capacidad de transmitir ese conocimiento

- El profesor/a contribuye decisivamente a la consecución de los objetivos / competencias esperados en la asignatura y demuestra un conocimiento reconocido en el campo además de la capacidad de transmitir ese conocimiento
- El profesor/a contribuye decisivamente a la consecución de los objetivos / competencias esperados en la asignatura; demuestra un conocimiento reconocido en el campo además de la capacidad de transmitir ese conocimiento y es capaz de aumentar el interés de los estudiantes en ese campo de formación
- El profesor/a contribuye decisivamente a la consecución de los objetivos / competencias esperados en la asignatura; demuestra un conocimiento reconocido en el campo además de la capacidad de transmitir ese conocimiento; es capaz de aumentar el interés de los estudiantes en ese campo de formación e influye en el desarrollo académico / profesional del estudiante

Breve CV de los autores

Luis Matosas-López

Profesor del Departamento de Economía Financiera, Contabilidad e Idioma Moderno, dentro del área de Informática Aplicada a las Ciencias Sociales en la Universidad Rey Juan Carlos. Sus trabajos de investigación se centran, por un lado, en la aplicabilidad de las TIC en el contexto universitario y, por otro, en el análisis de la calidad en la educación superior. Sobre ambos temas ha publicado distintos artículos científicos y capítulos de libro, realizando también múltiples comunicaciones en congresos nacionales e internacionales. En la actualidad es miembro del grupo de investigación CIBERIMAGINARIO de la Universidad Rey Juan Carlos, colaborador externo del grupo EDUTOOLS de la Universidad de León y miembro del colectivo internacional de docentes INNOVAGOGIA. ORCID ID: <https://orcid.org/0000-0001-7313-0146>. Email: luis.matosas@urjc.es

Alberto Romero-Ania

Doctor Europeo en Economía y desde 2006 profesor del Departamento de Economía Aplicada de la Universidad Rey Juan Carlos. Ha realizado una treintena de misiones docentes y/o investigadoras en: Unión Internacional de Telecomunicaciones (Naciones Unidas, Ginebra), IMC University of Applied Sciences Krems (Austria), University of Qingdao (China), University of Hué (Vietnam), Instituto de Turismo de Azerbaiyán (Baku), Hanoi University (Vietnam) y University of North Florida (Estados Unidos). Entre sus publicaciones destacan artículos en revistas internacionales del JCR (Journal Citation Reports) o varios libros publicados por Pearson. Asimismo, ha desarrollado tareas de técnico de emprendimiento y director de diferentes viveros de empresas del Ayuntamiento de Madrid. ORCID ID: <https://orcid.org/0000-0001-7865-9372>. Email: alberto.romero@urjc.es

Elena Cuevas-Molano

Profesora de la Universidad Rey Juan Carlos desde el 2006. Imparte docencia dentro del departamento de Ciencias de la Comunicación y Sociología en los grados de Publicidad y RR.PP., Diseño Integral y Gestión de la Imagen; Protocolo, Organización de Eventos y Comunicación Corporativa. Ha desempeñado perfiles de responsable en departamentos de

Publicidad, Imagen de Marca, y Marketing dentro de los sectores de medios de comunicación, telecomunicaciones y turismo, en entornos nacionales e internacionales. Actualmente centra su línea investigadora en la innovación docente aplicada a la educación universitaria y la estrategia corporativa. ORCID ID: <https://orcid.org/0000-0002-9876-8207>. Email: elena.cuevas@urjc.es