

Validez de las encuestas de satisfacción de los estudiantes para evaluar la calidad docente: el caso de la UPCT (Cartagena)

Validity of student satisfaction surveys to assess teaching quality: the UPCT case study (Cartagena, Spain)

García Martín, A.  **García Martín, A.**

antonio.gmartin@upct.es

antonio.gmartin@upct.es

Universidad Politécnica de Cartagena (España)

Universidad Politécnica de Cartagena (Spain)

Montero Cases, T.

teresa.montero@upct.es

teresa.montero@upct.es

Universidad Politécnica de Cartagena (España)

Universidad Politécnica de Cartagena (Spain)

García León, J.  **García León, J.**

josefina.leon@upct.es

josefina.leon@upct.es

Universidad Politécnica de Cartagena (España)

Universidad Politécnica de Cartagena (Spain)

Vázquez Arenas, G.  **Vázquez Arenas, G.**

gemma.vazquez@upct.es

gemma.vazquez@upct.es

Universidad Politécnica de Cartagena (España)

Universidad Politécnica de Cartagena (Spain)

Resumen

El programa DOCENTIA de ANECA se desarrolló como base para los modelos mediante los cuales las universidades españolas evalúan la docencia del profesorado,

Abstract

ANECA's DOCENTIA program was developed to support the evaluation of teaching by the Spanish universities. Nowadays, teaching quality evaluation

una necesidad ligada a los requerimientos de transparencia y de mejora continua que están en la esencia del EEES. Entre los indicadores habituales en estos modelos destacan las encuestas de satisfacción de los estudiantes, cuya validez es objeto de debate desde el origen de su utilización. Una de las opiniones críticas más extendidas sostiene que las encuestas no son una buena referencia, porque contienen sesgos inasumibles. Con el objetivo de comprobar si existen realmente esos sesgos, se ha realizado un estudio estadístico a partir de los datos de indicadores de calidad en la Universidad Politécnica de Cartagena. Se ha empleado la regresión lineal múltiple, mediante el procedimiento conocido como regresión hacia atrás, para identificar las variables que pudieran influir en las valoraciones que hacen los estudiantes y cuantificar su efecto. El principal resultado del estudio es que las variables analizadas, que suelen citarse como causas de sesgo, muestran una influencia poco o nada significativa sobre los resultados de las encuestas. Por otra parte, es importante subrayar que hay aspectos de la actividad docente que solamente pueden evaluarse a partir de la observación sistemática que hacen los estudiantes. Por todo ello consideramos que las encuestas de satisfacción son un indicador indispensable, y es fundamental que el peso que se les asigna en el modelo de evaluación docente refleje su importancia.

Palabras clave: calidad docente, evaluación del profesorado, DOCENTIA, validez, fiabilidad, regresión hacia atrás.

is a need linked to the transparency and continuous improvement requirements inherent to the European Higher Education Area. The indicators on the quality models frequently include student evaluation of teaching (SET), whose validity is the subject of a broad debate from the beginning of its use. One of the most widespread critical opinions argues that student surveys are not a good reference, as they contain non-assumable biases. With the aim of verify whether these biases actually exist, a statistical study has been carried out based on quality indicators data from the Technical University of Cartagena. Multiple linear regression has been used, through backward stepwise regression, to identify variables that may influence students' opinions and to quantify their effects. The main result is that the variables analyzed, many of which are often cited as causes of bias, show little or no significant influence on student satisfaction with teaching. It is also important to remark that there are aspects of the teaching activity that can only be evaluated from the systematic observation made by students. Therefore, we believe that student surveys are an essential indicator, and their weight within teaching evaluation models should reflect their relevance.

Key words: teaching quality, faculty evaluation, DOCENTIA, validity, reliability, backward stepwise regression.

Introducción

Cualquier proceso de mejora de la calidad requiere que se evalúe aquello que se pretende mejorar, y la docencia universitaria no puede ser una excepción a la regla. La transparencia en el funcionamiento de la universidad española, requisito esencial

del Espacio Europeo de Educación Superior, supone rendir cuentas a la sociedad que contribuye a financiarla (López-Aguado, 2018), y eso implica evaluar periódicamente la calidad de sus títulos, de su investigación y de su docencia, entre otros. La evaluación de la calidad docente se justifica, por tanto, por estas dos razones: la rendición de cuentas y la mejora continua. Si la evaluación de la docencia tiene como único fin mejorar la praxis docente, aportando información concreta al profesorado evaluado, es habitual hablar de evaluación docente formativa; cuando la emplean los gestores universitarios como base para tomar decisiones respecto a la promoción, contratación, etc. del profesorado, suele denominarse evaluación docente sumativa (Ramírez y Montoya, 2014).

La Agencia Nacional de Evaluación de la Calidad y Acreditación (ANECA) desarrolló el programa DOCENTIA de evaluación del profesorado, que sirve de base para que las universidades y las Agencias autonómicas diseñen sus modelos de evaluación, combinando información de distintas fuentes, organizada en dimensiones y criterios a los que se asignan pesos para obtener la valoración final. Entre las fuentes de información más usadas están las encuestas de satisfacción de los estudiantes con la actividad docente. Las encuestas de los estudiantes (*student evaluation of teaching*, SET) empezaron a emplearse, hace aproximadamente un siglo, en universidades anglosajonas, aunque su uso en España se ha generalizado bastante más tarde, a partir de los años 1980 (Casero, 2010). Los cuestionarios emplean, habitualmente, escalas tipo Likert y recogen la opinión de los estudiantes sobre distintos aspectos de la actividad docente. Autores como Marsh y Roche (2007) o Spooren *et al.* (2012) emplean el término *validity* (validez) para referirse a la capacidad de las SET para medir la calidad docente. Marsh y Roche (2007) emplean además el término *reliability* (fiabilidad).

El diseño de cada cuestionario depende de la Universidad que lo aplica, pero todos suelen incluir un ítem final, con el que los estudiantes valoran de manera global la labor del profesor, que a menudo se utiliza como valor medio representativo de cada encuesta. Las encuestas aportan información detallada de la labor docente de cada profesor en un formato muy adecuado para los modelos de evaluación de la docencia. Sin embargo, su uso es muy cuestionado, hasta el punto de haberse convertido en uno de los temas recurrentes de las publicaciones sobre educación universitaria. Entre los principales inconvenientes de las encuestas destacarían los siguientes:

1. Según algunos autores, las encuestas no miden la calidad docente sino otras cosas, por ejemplo, la popularidad del profesor (Emery *et al.*, 2003). En consecuencia, no existe según ellos relación entre lo que los estudiantes aprenden de un profesor y la valoración que hacen de su docencia (Uttl *et al.*, 2017).
2. Según otros autores, los resultados de las encuestas están sujetos a sesgos (Franklin, 2016). Entre esos sesgos estarían la carga de trabajo de los estudiantes y la indulgencia en las calificaciones (Greenwald y Gillmore, 1997), la representatividad de la muestra (Wolbring y Treischl, 2016), el curso en el que esta se sitúa, el tamaño del grupo, el género del profesorado (Boring, 2017), etc.

También son muchos los autores que se muestran favorables al uso de las encuestas, por entender que son una herramienta fiable y adecuada para evaluar la calidad docente (Marsh, 2007; Benton y Cashin, 2013; Hammonds *et al.*, 2016). En general, estos autores opinan que los posibles sesgos no afectan significativamente a la validez de las encuestas, y que estas, siempre que estén bien diseñadas, evalúan

aspectos de la docencia y no la popularidad del profesor. Centra (2003) advierte sobre el uso del concepto de sesgo que hacen algunos estudios, y sostiene que solo puede considerarse sesgo aquello que influye en la valoración sin tener relación con la calidad docente. Para este autor, la posible influencia de un grupo de clase numeroso, que dificulta el contacto con el estudiante y el uso de determinadas técnicas docentes, o de una carga de trabajo excesiva, que puede afectar al enfoque del aprendizaje y afectar a la coordinación docente, no deberían considerarse sesgos, ya que es lógico que estén relacionados con la calidad de la docencia.

Del mismo modo, Cohen (1983) se refiere a ciertas revisiones bibliográficas hechas con un reducido número de referencias elegidas de forma poco objetiva, y cuestiona el que se identifique la calidad docente con los resultados académicos altos, ya que tasas elevadas de éxito y rendimiento y altas calificaciones no implican docencia de calidad. Frente a los autores que afirman que las encuestas solo miden la popularidad del profesor, otros alegan que más bien evalúan su actitud respecto a la docencia, entre otros aspectos de esta, y que esa actitud es esencial para que el estudiante aprenda de manera significativa y constituye uno de los componentes de la eficacia como docente (Jones, 1989).

Sin embargo, incluso los más fervientes defensores de las encuestas se muestran partidarios de que la información que estas aportan, especialmente en los casos de evaluación docente sumativa, se complete con la de otras fuentes (Marsh, 2007). Entre esas fuentes se citan los informes de responsables académicos del Centro y el Departamento a los que está asignado el profesor evaluado, los resultados académicos de sus estudiantes (en ocasiones, contextualizados según el curso y el tipo de asignatura), el análisis de sus materiales docentes por parte de algunos colegas, la asistencia de otros docentes a sus clases e, incluso, la investigación desarrollada por él.

Por otra parte, Isla-Díaz *et al.* (2018) comentan la escasa discriminación de algunos modelos de DOCENTIA, que, posiblemente, introducen criterios poco adecuados con pesos que no parecen bien establecidos, y dan como resultado un exceso de valoraciones buenas o excelentes y muy pocas valoraciones mediocres o malas. Los informes de los responsables académicos suelen contribuir a que se obtengan altas valoraciones (Sánchez *et al.* 2009).

La situación es, por tanto, la siguiente: las encuestas de satisfacción suelen hacerse de forma sistemática y constituyen la herramienta más fácil de usar para evaluar la calidad docente, pero, un siglo después de que empezaran a utilizarse, el debate en torno a ellas sigue abierto. Las dudas sobre la validez de sus resultados, alimentadas por la discusión interminable que se recoge en la literatura sobre este tema, pueden llevar a los responsables académicos a minimizar el peso de las encuestas en el modelo de evaluación docente, bien sea por prudencia o bien porque se aprovecha ese debate para encubrir una falta de interés en institucionalizar su uso (Casero 2010).

Esta es la justificación del presente trabajo. Puesto que la validez de las encuestas va a depender, en buena medida, de la existencia o no de sesgos que pudieran poner en duda sus resultados, analizamos la influencia de varios factores clave en la satisfacción que muestran los estudiantes respecto a la docencia recibida. Si se comprueba que alguno de esos factores puede ser causa de un sesgo importante en las valoraciones, no cabe otra actuación que poner en duda la validez de las encuestas y limitar su peso, o

eliminarlas. Pero, si ocurre lo contrario, no estaría justificado hacerlo, o al menos no lo estaría por esa causa.

Objetivos

El principal objetivo de este trabajo es aportar criterios a la hora de asignar peso a las encuestas de los estudiantes en un modelo de evaluación de la docencia. Para ello se analizan las relaciones entre los valores de diversos indicadores de calidad (satisfacción de los estudiantes con la actividad docente, carga de trabajo percibida, resultados académicos, características de los grupos docentes y del profesorado, etc.) en busca de posibles sesgos.

Adicionalmente, se pretende con este trabajo contribuir a mejorar los sistemas de gestión de la información de la Universidad Politécnica de Cartagena (UPCT) desde la perspectiva de los procedimientos internos y externos de aseguramiento de la calidad.

Metodología

Se han utilizado los datos de indicadores de calidad de la UPCT durante el curso 2016-17, que incluyen satisfacción con la actividad docente, resultados académicos (calificación media, tasas de presentados, éxito y rendimiento), carga de trabajo percibida por los estudiantes, características de cada asignatura y profesor, etc. Se ha intentado tener en cuenta en el estudio aquellas variables, como la carga de trabajo o el sexo del profesor, que aparecen citadas como posibles causas de sesgo en la abundante bibliografía existente. Otras, como la raza del profesor, no tienen mucho sentido en nuestro contexto y, además, no disponemos de información para incluirlas en el análisis.

Los datos empleados son los que se recogen habitualmente en el marco de los sistemas de calidad de la Universidad, y no se ha realizado ninguna nueva encuesta *exprofeso*. Por eso no han introducido en el estudio variables como la percepción de los estudiantes sobre la dificultad de cada asignatura, pero sí otras como los resultados académicos obtenidos, que están relacionados con la dificultad. La carga de trabajo percibida se mide en la UPCT mediante una encuesta cuatrimestral, tal como se explica más adelante.

Los datos proceden de 16 títulos de Grado y 21 de Máster. Entre los de Grado, 14 pertenecen a la rama de Ingeniería y Arquitectura, mayoritaria en la UPCT, y 2 a la de Ciencias Sociales y Jurídicas. De los títulos de Máster, 6 son habilitantes, ya que conducen a otras tantas profesiones reguladas, todas ellas ingenierías.

La encuesta de satisfacción con la actividad docente (S) de la UPCT aplica una escala tipo Likert entre 1 (opción menos favorable) y 5 (opción más favorable). Desde el curso 2014-15 se emplea un cuestionario de 11 ítems, en el que se ha intentado recoger únicamente aquella información sobre la actividad docente del profesorado que los estudiantes pueden, razonablemente, aportar y de modo que sea suficiente para cubrir las necesidades de los procesos de seguimiento de los títulos (tabla 1).

Tabla 1. Los 11 ítems del cuestionario de la encuesta de satisfacción de los estudiantes con la actividad docente (S) de la UPCT.

1	La actitud del profesor ha fomentado mi interés por la asignatura
2	Las explicaciones del profesor son claras
3	Los materiales didácticos (apuntes, vídeos, contenidos del aula virtual, etc.) que emplea el profesor son útiles para mi aprendizaje
4	El profesor fomenta la participación del alumnado
5	El profesor se muestra accesible para la realización de consultas sobre la asignatura fuera de clase
6	Los resultados que proporciona el profesor sobre las actividades de evaluación continua permiten conocer mi progreso durante el curso
7	Las actividades encargadas por el profesor en horario no presencial son útiles para mi aprendizaje
8	La metodología docente utilizada me ayuda a desarrollar los conocimientos y habilidades objetivo de la asignatura
9	La metodología de evaluación utilizada me parece adecuada para valorar la adquisición de conocimientos y habilidades en la asignatura
10	Estoy adquiriendo los conocimientos y habilidades objetivo de la asignatura
11	En términos generales, estoy satisfecho con la actividad docente desarrollada por el profesor

La consistencia interna del cuestionario se analizó mediante el coeficiente alfa de Cronbach, obteniéndose un valor muy alto (0,98). Por tanto, en lo sucesivo nos referiremos solo al ítem global de la encuesta, ítem 11 (“En términos generales, estoy satisfecho con la actividad docente desarrollada por el profesor”), como representativo del conjunto, aunque en el estudio se han tenido en cuenta todos ellos. La encuesta, común a todos los títulos de la Universidad, se pasa para cada profesor de cada asignatura/grupo, en asignaturas obligatorias y optativas de títulos oficiales de Grado y Máster. En los títulos de Grado y en los de Máster habilitante se rellena de forma presencial, mientras que en los de Máster no habilitante se hace *on-line*.

La carga de trabajo percibida (CT) también se cuantifica con una escala tipo Likert 1-5, pero en este caso el valor 1 de la escala se refiere a una carga muy baja y el 5 a una carga excesiva, siempre con relación a la carga docente asociada a la asignatura y según la percepción de los estudiantes que la cursan (García Martín y García-León, 2017). Por lo tanto, el valor medio de la escala, 3, corresponde a asignaturas cuya carga de trabajo coincide con la carga docente prevista en la memoria de verificación del título. La encuesta de carga de trabajo percibida se pasa para cada grupo de cada asignatura, independientemente del número de profesores que lo imparten, y su objetivo es identificar aquellas asignaturas cuya carga de trabajo real está descompensada y puede dificultar la coordinación horizontal del curso. Solo se hace en las asignaturas obligatorias, no en las optativas, en títulos de Grado y en los de Máster que sean habilitantes.

La base de datos empleada para este trabajo procede de la Oficina de Prospección y Análisis de Datos (OPADA) de la UPCT y, además de los datos de las encuestas de satisfacción con la actividad docente y de carga de trabajo percibida, contiene la información correspondiente a:

- características de la asignatura: centro, título, curso, grupo, código, etc.
- resultados académicos del grupo: número de matriculados, calificación media, tasas de presentados, éxito y rendimiento, etc.
 - la tasa de presentados (TP) es el porcentaje de estudiantes presentados a examen respecto a matriculados.
 - la tasa de éxito (TE) es el porcentaje de estudiantes aprobados respecto a presentados a examen.
 - la tasa de rendimiento (TR) es el porcentaje de estudiantes aprobados respecto a matriculados.
- datos del profesorado: sexo, categoría académica, antigüedad en la UPCT, etc.

Se dispone, además, del número de respuestas a cada encuesta, tanto en la de satisfacción como en la de carga de trabajo.

La base de datos consta de un total de 1.592 registros. En algunos de ellos no se dispone de datos de satisfacción, por lo que los más relevantes para nuestro estudio son los 1.561 registros que corresponden al conjunto de encuestas de satisfacción con la actividad docente realizadas durante el curso 2016-17 en la UPCT. Casi todos los registros disponen de los resultados académicos de la asignatura/grupo correspondiente. Faltan datos de profesorado en algunas de las asignaturas impartidas en centros adscritos, de forma que los datos completos del profesorado están disponibles en 1.442 de los registros. Por otra parte, solo se dispone de datos de carga de trabajo percibida en 1.318 registros, que son los que corresponden a asignaturas obligatorias de títulos de Grado y de Máster habilitante.

El análisis se ha hecho tanto para el conjunto de datos disponibles como discriminando por niveles, tipos de asignaturas, cursos, centros, títulos, ramas, categoría académica del profesorado, género de este, etc. Se han comparado entre sí los resultados obtenidos, en busca de diferencias que pudieran limitar su uso como indicadores del desempeño docente. Se ha estudiado la correlación entre variables, particularmente entre la satisfacción con la actividad docente y las restantes, calculando los coeficientes de correlación y trazando los gráficos de dispersión correspondientes.

Para estudiar la posible existencia de sesgos en las encuestas de satisfacción con la actividad docente se ha realizado un análisis de regresión lineal múltiple, partiendo de aquellas variables contenidas en la base de datos que puedan identificarse con las que suelen aparecer en publicaciones sobre las encuestas. Pretendemos localizar las variables que influyen de forma significativa en la satisfacción y cuantificar su influencia, para disponer de argumentos a la hora de discutir la pertinencia del uso de los resultados de las encuestas en los programas de evaluación de la calidad docente.

Puesto que una de las variables más relevantes a priori era la carga de trabajo (CT), solo se han considerado en el análisis de regresión lineal múltiple aquellos registros de la base de datos que disponen de esa información. Además, se ha optado por eliminar de la

muestra los registros que tienen menos de 10 respuestas en la encuesta de satisfacción con la actividad docente, lo que deja un total de 975 observaciones.

Hemos planteado varios modelos de regresión lineal múltiple para evaluar el impacto en la variable satisfacción de distintas combinaciones de las variables de interés. En todos los casos la variable dependiente o respuesta ha sido la valoración media obtenida en las encuestas de satisfacción con la actividad docente por un profesor en un grupo, es decir la valoración del ítem 11 de las encuestas. Empleamos un proceso iterativo para obtener el modelo reducido de influencia, procedimiento conocido como regresión hacia atrás (*Backward Stepwise Regression*). En este procedimiento se parte de un modelo completo, que incorpora todos los efectos que pueden influir en la respuesta, y en cada etapa se elimina la variable menos influyente, esto es, la variable con el p-valor más alto. Una vez eliminada esta variable, se vuelve a estimar el modelo y se procede de la misma manera que en la estimación anterior. Este proceso se continúa hasta que todas las variables incluidas en el modelo reducido tengan coeficientes significativamente distintos de cero.

Tras comprobar con modelos anteriores que la carga de trabajo (CT) presenta cierta influencia en la satisfacción (S), se ha intentado delimitar su efecto e identificar cómo afectan sus valores extremos, por exceso o por defecto; para ello se ha considerado la carga de trabajo por intervalos:

- carga de trabajo baja (CT1) es una variable binaria que vale 1 si $(CT) < 3$ y 0 en otro caso
- carga de trabajo media (CT2) es una variable binaria que vale 1 si $la (CT) \in [3,4)$ y 0 en otro caso
- carga de trabajo alta (CT3) es una variable binaria que vale 1 si $la (CT) \in [4,5]$ y 0 en otro caso

Además, se han generado tres nuevas variables para poder introducir en el estudio otros aspectos de interés: expectativa (EXP) es la diferencia entre las tasas de éxito y rendimiento; género (GEN) es una variable binaria, cuyo valor depende de si el profesor es hombre o mujer; categoría académica (CA) es una variable binaria que depende de si el profesor pertenece o no a los cuerpos docentes universitarios.

Resultados

La tabla 2 resume el estudio estadístico descriptivo realizado sobre la base de datos. Aunque se han tenido en cuenta muchas otras variables, solo se muestran en la tabla las que consideramos más significativas. Se han empleado todos los datos disponibles de cada una de esas variables, incluyendo los registros a los que les falta alguna información, y sin tener en cuenta el número de respuestas a las encuestas.

El valor medio en la UPCT del ítem global 11 de la encuesta de satisfacción con la actividad docente (S) es de 3,72 en la escala Likert 1-5. El 14,1 % de las encuestas no alcanza el valor 3 de la escala. Un 43 % de las encuestas obtienen valores iguales o superiores a 4, y las encuestas con resultados excelentes, entre 4,5 y 5, suponen un

13,5 % del total. Se observa que hay encuestas que han alcanzado el valor máximo de la escala, pero también las hay que han obtenido el mínimo. El número de respuestas de la encuesta de satisfacción varía entre 3 y 56 estudiantes.

Las asignaturas optativas obtienen, en media, valores de satisfacción (S) algo más altos que las obligatorias. Si nos centramos en los títulos de Grado y solo consideramos las asignaturas obligatorias, se obtiene que la satisfacción media es muy constante a lo largo de los cursos, oscilando entre 3,63 (segundo curso) y 3,67 (cuarto curso). Por otra parte, los títulos de Máster obtienen una satisfacción media (3,85) algo superior a la de los títulos de Grado (3,70).

Tabla 2. Estadísticos descriptivos de la satisfacción de los estudiantes con la actividad docente (S) y de las otras variables consideradas en el estudio. Datos de la UPCT, curso 2016-17.

Variable	Media	Muestra	Varianza	Máximo	Mínimo
Satisfacción (S): ítem 11	3,72	1.561	0,53	5,00	1,00
Encuesta satisfacción (S): muestra	15,13	1.561	85,46	56	3
Carga de trabajo percibida (CT)	3,45	1.318	0,38	5,00	1,71
Encuesta carga de trabajo (CT): muestra	16,22	1.318	94,89	52	1
Tamaño de grupo (TG)	30,13	1.592	389,35	104	1
Calificación media (CM)	6,20	1.587	1,90	10,00	2,43
Tasa de presentados (TP)	84,34	1.592	248,43	100,00	0,00
Tasa de éxito (TE)	83,58	1.587	324,06	100,00	0,00
Tasa de rendimiento (TR)	71,96	1.592	566,94	100,00	0,00

Fuente: elaboración propia.

De los 1.442 registros en los que existe información sobre las características del profesorado objeto de la encuesta, solo un 23,7 % corresponde a profesoras. El porcentaje de profesoras varía mucho de unos centros a otros, siendo máximo en el que oferta títulos de la rama de Ciencias Sociales y Jurídicas y mínimo en los de determinadas ingenierías, situación que resulta habitual en la universidad española (Elizondo *et al.* 2010). El porcentaje de alumnas, 29,7 %, es mayor que el de profesoras, pero muy inferior a la media de la universidad española, como suele ocurrir en las enseñanzas técnicas. El valor medio de satisfacción obtenido por las profesoras es de 3,73, mientras que el de profesores varones es prácticamente idéntico, 3,74.

En cuanto a la carga de trabajo percibida (CT), su valor medio, 3,45, es superior al valor central de la escala Likert empleada, lo que significa que la carga de trabajo es, en media y según la percepción de los estudiantes, algo más alta que la prevista en los planes de estudios. Esto ocurre de manera habitual desde que se cuantifica esta variable, y se comprueba que los títulos de Máster habilitante suelen ser los que mayores valores medios presentan. Como se ha indicado, la carga de trabajo no se mide en las asignaturas optativas ni en las obligatorias de los títulos de Máster no habilitante.

La calificación media (CM) obtenida por los estudiantes oscila entre los valores 2,43 y 10. Los valores extremos, máximos y mínimos, de este tipo de variables suelen corresponder a asignaturas con muy pocos estudiantes matriculados. Lo mismo ocurre con las tasas de presentados (TP), éxito (TE) y rendimiento (TR).

La Tabla 3 muestra los coeficientes de correlación entre la variable satisfacción con la actividad docente (S) y otras variables de interés, así como el resultado de los test de correlación de Pearson y de Spearman (que es una técnica no paramétrica que se utiliza cuando no se dan los supuestos de normalidad). También aquí se han usado todos los datos disponibles, sin eliminar aquellos cuya muestra es pequeña. La antigüedad del profesorado se ha calculado desde la creación de la UPCT, hace poco más de 20 años, ya que no se dispone de datos de la época anterior, en la que la UPCT formaba parte de la Universidad de Murcia.

Tabla 3. Coeficientes de correlación de algunas de las variables analizadas con la satisfacción de los estudiantes con la actividad docente (S). Datos de la UPCT, curso 2016-17.

	Correlación con satisfacción S	p-valor
Carga de trabajo percibida (CT)	-0,12	<0,01
Tamaño de grupo (TG)	-0,18	<0,01
Calificación media de los estudiantes (CM)	0,22	<0,01
Tasa de presentados (TP)	0,15	<0,01
Tasa de éxito (TE)	0,18	<0,01
Tasa de rendimiento (TR)	0,19	<0,01
Antigüedad del profesor en la UPCT	-0,06	<0,05

Fuente: elaboración propia.

Todos los coeficientes de correlación de la tabla 3 muestran valores débiles o muy débiles y su signo es el esperado, salvo en el caso de la antigüedad. Sin embargo, las correlaciones entre la satisfacción S y las variables consideradas resultan ser estadísticamente significativas, tanto con el test de correlación de Pearson como con el test de rangos de Spearman, lo que significa que la influencia de dichas variables no puede considerarse nula.

La Figura 1 muestra el gráfico de dispersión de la variable satisfacción (S) y la carga de trabajo percibida (CT). Se observa que es posible obtener valores altos de satisfacción incluso con cargas de trabajo elevadas.

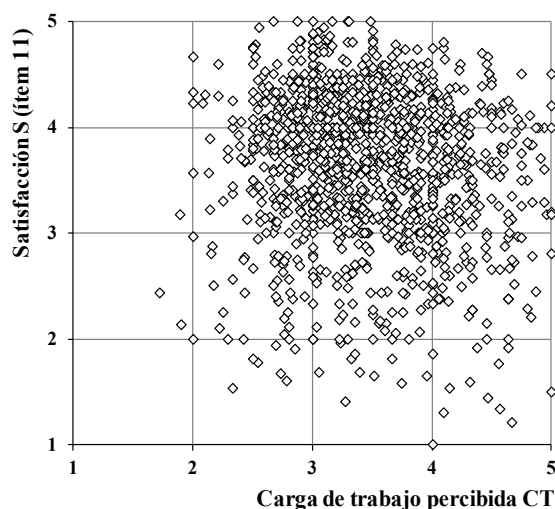


Figura 1. Gráfico de dispersión entre la satisfacción de los estudiantes con la actividad docente (S) y la carga de trabajo percibida por los estudiantes (CT). Datos de la UPCT, curso 2016-17. Fuente: elaboración propia.

En el análisis de regresión lineal múltiple se han ensayado distintas combinaciones de variables mediante el procedimiento anteriormente descrito. Los resultados de todas ellas han sido similares a los que se muestran, a modo de ejemplo, en la tabla 4. En este caso, las variables consideradas han sido la calificación media de los estudiantes (CM), la carga de trabajo percibida (CT), el tamaño de grupo (TG), la variable expectativa (EXP), la tasa de presentados (TP), el género del profesor (GEN) y su categoría académica (CA).

Tabla 4. Resultados de la regresión lineal múltiple considerando la carga de trabajo (CT) por intervalos. Variable dependiente: satisfacción con la actividad docente (S). Solo se han usado registros con 10 o más respuestas en la encuesta de satisfacción.

Variable	Modelo Completo		Modelo Reducido	
	Coefficientes	p-valor	Coefficientes	p-valor
Cte.	2,7779143	3,45e-10 ***	2,785271	6,43e-13 ***
CM	0,0179711	0,5632	-	-
CT2	0,0188061	0,7484	-	-
CT3	-0,1510429	0,0299 *	-0,184078	0,000496 ***
TG	0,0004899	0,7125	-	-
EXP	0,0096804	0,1496	0,011740	0,033454 *
TP	0,0080305	0,1235	0,008873	0,021232 *
GEN	-0,0055676	0,9193	-	-
CA	-0,0303568	0,5400	-	-
F-statistic: 2,153 con 8 y 846 DF y con p-valor = 0,02899			F-statistic: 5,988 con 3 y 959 DF p-valor = 0,0004819	

(***) sig. al 99,9 %, (*) sig. al 95 % y (-) sig. al 90 %.
Fuente: elaboración propia

El modelo reducido de la tabla 4, representativo de los obtenidos partiendo de otros modelos completos, muestra las variables que tienen una influencia estadísticamente significativa sobre la satisfacción de los estudiantes con la actividad docente. Esta influencia resulta ser, en todos los casos en los que existe, débil, y suele producirse en la dirección esperada. El estudio se ha replicado empleando solo los datos de títulos de Grado, y no los de Máster, con idénticos resultados.

La carga de trabajo percibida (CT) muestra una influencia negativa débil sobre la satisfacción, pero se observa que solo los valores altos/excesivos de carga de trabajo (CT3) presentan un efecto estadísticamente significativo respecto a los valores bajos (CT1), que es el intervalo omitido en la regresión, mientras que el intervalo de carga de trabajo media (CT2) no es significativo con respecto a los valores bajos (CT1).

Las variables tasa de presentados a examen (TP) y expectativa (EXP) muestran una influencia positiva muy débil sobre la satisfacción.

Las variables tamaño de grupo (TG) y categoría académica (CA) no influyen significativamente en la satisfacción (S). El tamaño de grupo, además de como variable cuantitativa, se incluyó en el análisis por intervalos (grupos grandes, con más de 50 matriculados, y pequeños, con un número de matriculados menor o igual a 50), pero tampoco así se observa relación con la satisfacción. La categoría académica del profesorado (CA) se ha incluido de dos formas distintas: por una parte, hemos considerado

si el profesor es un trabajador con dedicación exclusiva en la UPCT o un trabajador a tiempo parcial; por otra parte, como se indicó, se ha considerado si el profesor pertenece a los cuerpos docentes universitarios o bien es un profesor con contrato laboral. En ambos casos, la categoría del profesorado no muestra influencia significativa sobre la satisfacción del estudiante con la actividad docente.

La calificación media obtenida por los estudiantes (CM) y el género del profesorado (GEN) no muestran una influencia estadísticamente significativa en la satisfacción (S).

Los valores de R^2 de todos estos modelos están en el rango del 10 % al 2 %, es decir que la proporción de la varianza total explicada por dichos modelos es muy reducida.

Conclusiones y discusión

En este trabajo se ha estudiado la influencia de distintas variables en la satisfacción de los estudiantes de la UPCT con la actividad docente. Se aportan así criterios para decidir cuál debe ser el peso de las encuestas de satisfacción en un modelo de evaluación de la docencia, que es el principal objetivo del trabajo.

Se ha comprobado que los valores medios de satisfacción de las asignaturas optativas son algo mayores que los de las asignaturas obligatorias. Este resultado es lógico, puesto que las optativas ya han sufrido un proceso de selección, de manera que aquellas que son elegidas por pocos estudiantes acaban por no ofertarse, pero conviene tenerlo en cuenta a la hora de incluir unas y otras en la evaluación de la actividad docente. No se han detectado diferencias significativas entre los distintos cursos de los títulos de Grado, pero sí entre los niveles de Grado y Máster.

Si se sitúa en valores reducidos o medios, la carga de trabajo percibida no influye de forma significativa en la satisfacción, mientras que las cargas de trabajo elevadas sí lo hacen. Esta situación no puede interpretarse como un sesgo, ya que los valores extremos de carga de trabajo representan en realidad una inadecuada planificación docente, que afecta a la calidad docente y, consecuentemente, también a la satisfacción (Centra, 2003). En cualquier caso, la influencia de la carga de trabajo, incluso limitada al intervalo de carga elevada, es débil.

Las variables tasa de presentados (TP) y expectativa (EXP) muestran una influencia positiva sobre la satisfacción, pero esa influencia es muy débil. Fernández Rico *et al.* (2007) sugieren emplear las expectativas de los estudiantes previas al examen final como variable explicativa, y encuentran que tiene un efecto significativo en la satisfacción.

Es importante destacar que ni la calificación media de los estudiantes ni el género del profesorado influyen de manera significativa en la satisfacción. Cuando rellenan las encuestas de satisfacción los estudiantes no saben cuál será su calificación final, pero sí conocen algunos resultados parciales, además de los resultados obtenidos por sus compañeros de cursos anteriores. A pesar de ello, el análisis realizado no muestra que exista un sesgo por indulgencia, ya que no hay una relación significativa entre calificaciones altas y buenas valoraciones en la encuesta de satisfacción. El hecho de que no exista sesgo de género es más relevante aún, ya que hay varios estudios que sostienen

lo contrario (Boring, 2017) y la supuesta presencia de ese sesgo ha sido utilizada como argumento de peso por muchos autores.

Este estudio proporciona, por lo tanto, una evidencia de ausencia de relación lineal entre la satisfacción con la actividad docente y las variables que habitualmente suelen citarse como causa de sesgo. Ese era el resultado lógico si se sostiene la tesis de no existencia de sesgos en las encuestas, incluso habiendo encontrado pequeñas correlaciones entre esas variables y la satisfacción.

Somos conscientes de que el estudio emplea datos de una sola Universidad, pero sus resultados coinciden con los de autores como Marsh (2007) o Centra (2003) y nuestras conclusiones no pueden ser diferentes de las de ellos.

Por otra parte, autores como Abadía *et al.* (2015) han estudiado cuáles son las características que debería tener un buen docente, desde distintas perspectivas. De la revisión de la extensa bibliografía existente sobre este tema, se concluye que la calidad docente implica, al menos:

- cumplir las obligaciones docentes: respetar el horario de tutorías, los plazos de las convocatorias de examen y de la entrega de calificaciones, etc.
- una buena planificación docente: carga de trabajo equilibrada y bien distribuida a lo largo del cuatrimestre, actividades docentes pertinentes y útiles, evaluación adaptada a lo establecido en la guía docente, etc.
- una actitud correcta: disponibilidad, interés por la docencia, interacción con los estudiantes, capacidad de crear un clima idóneo para el aprendizaje, etc.
- habilidades comunicativas: capacidad para transmitir adecuadamente ideas y conocimientos.

Cualquier estudiante que asista a clase con cierta regularidad es capaz de evaluar, con criterio suficiente, todos estos aspectos de la actividad docente de un determinado profesor, además de otros que no se citan aquí. Es más, para algunos de dichos aspectos, la observación sistemática que hacen los estudiantes que asisten a clase es la única forma de asegurar una evaluación objetiva. El hecho de que los resultados de las encuestas sean valores estadísticos para los que se exige una cierta muestra, garantiza que, al menos en la mayoría de los casos, las opiniones extremas van a tener poca incidencia.

Todas las habilidades que requiere la buena docencia se pueden desarrollar y aprender, pero eso requiere disponer de tiempo. El profesorado universitario tiene que dedicarse a la investigación y, a menudo, también a la gestión, y en ambos casos dispone de indicadores suficientemente establecidos para evaluar su labor. Por lo tanto, la investigación y la gestión son ocupaciones que percibe como prioritarias, y más rentables, y a las que suele dedicar mucho tiempo. La docencia seguirá siendo la menos importante de sus tareas en tanto no se le dé el valor que merece (Mas Torelló, 2012), y, para ello, es fundamental evaluarla mediante herramientas apropiadas y otorgarle un peso significativo en los modelos de evaluación de la calidad docente.

Respecto a la conveniencia de completar la información de las encuestas con la de otras fuentes, existe consenso entre los autores consultados, entre ellos Marsh y Roche (1997), Murray (2005), Benton y Cashin (2013), Berk (2014) y Franklin (2016). No obstante, resulta complicado obtener información objetiva de la mayoría de esas fuentes: los resultados académicos no representan la multidimensionalidad del proceso de aprendizaje (Marsh y Roche, 1997); la evaluación por parte de otros profesores o de los responsables académicos puede depender de la relación que estos mantengan con el profesor evaluado; los materiales docentes a disposición de los estudiantes no son comparables y dependerán del tipo de docencia que se desarrolle y de algunas de las competencias asociadas a la asignatura -por ejemplo, en el caso de una asignatura que trabaje las competencias relacionadas con la búsqueda de información puede que no sea conveniente que disponga de apuntes completos-; la relación entre la producción científica de un profesor y la calidad de su docencia no está demostrada y nos llevaría a la paradoja de considerar que para mejorar la docencia hay que dedicarse menos a ella y más a la investigación.

La dificultad de disponer de información adicional a las encuestas de satisfacción de los estudiantes, que sea adecuada y pueda obtenerse de forma sistemática, hace que se corra el riesgo de introducir en la evaluación de la calidad docente indicadores que poco o nada tienen que ver con ella, y de diseñar modelos de evaluación con poca capacidad para detectar a los docentes que realmente destacan (Isla-Díaz *et al.*, 2018). Frente a indicadores que aportan pocos matices, los resultados de las encuestas de satisfacción se distribuyen de forma que la mayoría del profesorado obtiene valoraciones equivalentes al aprobado o al notable, mientras que son relativamente bajos los porcentajes de valoraciones sobresalientes. También lo son, por cierto, los de valoraciones que equivalen a un suspenso.

Por lo demás, y de acuerdo con autores como Marsh (2007), hay que insistir en la conveniencia de emplear para las encuestas un cuestionario bien diseñado, que solo incluya preguntas adaptadas a los estudiantes y a aquellas dimensiones de la docencia que ellos pueden observar y evaluar. Además, los resultados de las encuestas expresarán aún mejor la labor del profesor si se emplean valores promedio de varias asignaturas en distintos cursos académicos, incluso eliminando las valoraciones extremas, si las hubiese, que pudieran ser poco representativas.

La conclusión final de este trabajo es que las valoraciones de los estudiantes en las encuestas de satisfacción con la actividad docente no están influidas por factores como las calificaciones obtenidas, el tamaño de grupo o el género del profesor, y apenas lo están por otras como la carga de trabajo o las tasas de éxito y rendimiento. No se han detectado sesgos, como el de género, que puedan servir de argumento para cuestionar su validez en la evaluación de la calidad docente. En consecuencia, consideramos que los resultados de las encuestas de satisfacción, sin ser una herramienta perfecta, constituyen un buen indicador para medir la calidad docente, al que habría que dar un peso en cualquier modelo de evaluación no inferior a una tercera parte de la valoración total.

Referencias

- Abadía, A.R., Bueno, C., Ubieto-Artur, M.I., Márquez, M.D., Sabaté, S., Jorba, H., Pagès, T. (2015). Competencias del buen docente universitario. Opinión de los estudiantes. *REDU, Revista de Docencia Universitaria*, 13(2), 363-390. <https://doi.org/10.4995/redu.2015.5453>
- Benton, S.L., Cashin, W.E. (2013). Student ratings of instruction in college and university courses. En M.B. Paulsen (ed.), *Higher Education: Handbook of Theory and Research* (pp. 279-326). https://doi.org/10.1007/978-94-017-8005-6_7
- Berk, R.A. (2014). Should student outcomes be used to evaluate teaching? *Journal of Faculty Development*, 28(2), 87-96. Recuperado de <https://pdfs.semanticscholar.org/72dc/eb08a40f9019c72d5f20971bd3413fe06efd.pdf>
- 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>
- Casero, A. (2010). Factores moduladores de la percepción de la calidad docente. *RELIEVE, Revista Electrónica de Investigación y Evaluación Educativa*, 16(2), 1-17. <https://doi.org/10.7203/relieve.16.2.4135>
- Centra, J.A. (2003). Will teachers receive higher student evaluations by giving higher grades and less course work? *Research in Higher Education*, 44(5), 495-518. <https://doi.org/10.1023/A:1025492407752>
- Cohen, P.A. (1983). Comment on a selective review of the validity of student ratings of teaching. *The Journal of Higher Education*, 54(4), 448-458. <https://doi.org/10.2307/1981907>
- Elizondo, A., Novo, A., Silvestre, M. (2010). *Igualdad de mujeres y hombres en las universidades españolas*. Instituto de la Mujer, Madrid. Recuperado de <http://www.inmujer.gob.es/areasTematicas/estudios/serieEstudios/docs/igualdadUniversidades.pdf>
- Emery, C.R., Kramer, T.R., Tian, R.G. (2003). Return to academic standards: a critique of student evaluations of teaching effectiveness. *Quality Assurance in Education*, 11(1), 37-46. <https://doi.org/10.1108/09684880310462074>
- Fernández Rico, J.E., Fernández, S., Álvarez, A., Martínez Cambolor, P. (2007). Éxito académico y satisfacción de estudiantes con la enseñanza universitaria. *RELIEVE, Revista Electrónica de Investigación y Evaluación Educativa*, 13(2). <https://doi.org/10.7203/relieve.13.2.4207>
- Franklin, M. (2016). Student evaluations of teaching in business and accounting courses: a perspective and a suggested improvement. *Cogent Business & Management*, 3(1), 1-17. <https://doi.org/10.1080/23311975.2016.1226458>
- García Martín, A., García-León, J. (2017). Una experiencia de medición de la carga de trabajo percibida por los estudiantes para facilitar la coordinación horizontal. *REDU, Revista de Docencia Universitaria*, 15(1), 81-104. <https://doi.org/10.4995/redu.2017.5987>
- Greenwald, A.G., Gillmore, G.M. (1997). No pain, no gain? The importance of measuring course workload in student ratings of instruction. *Journal of Educational Psychology*, 89(4), 743-751. <https://doi.org/10.1037/0022-0663.89.4.743>

- Hammonds, F., Mariano, G.J., Ammons, G., Chambers, S. (2016). Student evaluations of teaching: improving teaching quality in higher education. *Perspectives: Policy and Practice in Higher Education*, 21(1), 26-33. <https://doi.org/10.1080/13603108.2016.1227388>
- Isla-Díaz, R., Marrero-Hernández, H., Hess-Medler, S., Soriano, M., Acosta-Rodríguez, S., Pérez-Monteverde, M.V., Blanco-Freijo, M. (2018). Una mirada longitudinal: ¿Es el “Docentia” útil para la evaluación del profesorado universitario? *RELIEVE, Revista Electrónica de Investigación y Evaluación Educativa*, 24(2). <https://doi.org/10.7203/relieve.24.2.12142>
- Jones, J. (1989). Students’ ratings of teacher personality and teaching competence. *Higher Education*, 18(5), 551-558. <https://doi.org/10.1007/BF00138747>
- López-Aguado, M. (2018). La evaluación de la calidad de títulos universitarios. Dificultades percibidas por los responsables de los sistemas de garantía de calidad. *Educación XX1*, 21(1), 263-284. <https://doi.org/10.5944/educxx1.20195>
- Marsh, H., Roche, L. (1997). Making students’ evaluations of teaching effectiveness effective. *American Psychologist*, 52(11), 1187-1197. <https://doi.org/10.1037/0003-066X.52.11.1187>
- Marsh, H. (2007). Students’ evaluations of university teaching: dimensionality, reliability, validity, potential biases and usefulness. En R.P. Perry and J.C. Smart (eds.), *The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective*, 319-383. Springer. https://doi.org/10.1007/1-4020-5742-3_9
- Mas Torelló, O. (2012). Las competencias del docente universitario: la percepción del alumno, de los expertos y del propio protagonista. *REDU - Revista de Docencia Universitaria*, 10(2), 299-318. <https://doi.org/10.4995/redu.2012.6109>
- Murray, H.G. (2005). *Student evaluation of teaching: has it made a difference?* Paper presented at the Annual Meeting of the Society for Teaching and Learning in Higher Education, Charlottetown, Prince Edward Island. Recuperado de <https://www.stlhe.ca/wp-content/uploads/2011/07/Student-Evaluation-of-Teaching1.pdf>
- Ramírez, M.I., Montoya, J. (2014). La evaluación de la calidad de la docencia en la universidad: Una revisión de la literatura. *REDU, Revista de Docencia Universitaria*, 12 (2), 77-95. <https://doi.org/10.4995/redu.2014.5641>
- Sánchez, F., Rubio, R., Alonso, E., Retamal, K. (2009). La valoración de la actividad docente. Algo más que la opinión de los estudiantes. *Boletín de Psicología*, 97, 71-92. Recuperado de <https://www.uv.es/seoane/boletin/previos/N97-5.pdf>
- Spooren, P., Mortelmans, D., Thijssen, P. (2012). ‘Content’ versus ‘style’: acquiescence in student evaluation of teaching? *British Educational Research Journal*, 38(1), 3-21. <https://doi.org/10.1080/01411926.2010.523453>
- Uttl, B., White, C.A., Gonzalez, D. (2017). Meta-analysis of faculty’s teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation* 54, 22-42. <https://doi.org/10.1016/j.stueduc.2016.08.007>
- Wolbring, T., y Treischl, E. (2016). Selection bias in students’ evaluation of teaching. Causes of student absenteeism and its consequences for course ratings and rankings. *Research in Higher Education*, 57, 51-71. <https://doi.org/10.1007/s11162-015-9378-7>