

2

LA CALIDAD DE LA DOCENCIA PERCIBIDA POR EL ALUMNADO Y EL PROFESORADO UNIVERSITARIOS: ANÁLISIS DE LA DIMENSIONALIDAD DE UN CUESTIONARIO DE EVALUACIÓN DOCENTE

(THE QUALITY OF TEACHING AS PERCEIVED BY STUDENTS AND UNIVERSITY TEACHERS: ANALYSIS OF THE DIMENSIONALITY OF A TEACHER EVALUATION QUESTIONNAIRE)

María Castro Morera
Enrique Navarro Asencio
Ángeles Blanco Blanco †
Universidad Complutense de Madrid

DOI: 10.5944/educXX1.25711

Cómo referenciar este artículo/How to reference this article:

Castro Morera, M.; Navarro Asencio, E. y Blanco Blanco, Á. (2020). La calidad de la docencia percibida por el alumnado y el profesorado universitarios: análisis de la dimensionalidad de un cuestionario de evaluación docente. *Educación XX1*, 23(2), 41-65, doi: 10.5944/educXX1.25711

Castro Morera, M.; Navarro Asencio, E. & Blanco Blanco, Á. (2020). The quality of teaching as perceived by students and university teachers: analysis of the dimensionality of a teacher evaluation questionnaire. *Educación XX1*, 23(2), 41-65, doi: 10.5944/educXX1.25711

RESUMEN

Tras 75 años de investigación sobre los cuestionarios de evaluación de los estudiantes sobre la enseñanza, se mantiene abierto el debate sobre su dimensionalidad. Esta cuestión tiene importantes repercusiones, tanto teóricas (para la validación del constructo “calidad de la docencia”) como prácticas (asociándose ventajas a las medidas unidimensionales para fines sumativos y a las multidimensionales para fines formativos).

La presente investigación se propone un análisis de la dimensionalidad de un cuestionario estándar de evaluación de la docencia por parte del alumnado; así como completar el análisis del constructo *calidad de la*

docencia añadiendo la perspectiva del profesorado, profundizando en la naturaleza de lo que unos y otros perciben como una docencia de calidad.

La doble perspectiva adoptada en este estudio se sustenta en el uso del mismo cuestionario de valoración de la actividad docente con los dos colectivos implicados, profesores y estudiantes de la Universidad Complutense de Madrid (UCM). Se analizan un total de 125.155 encuestas cumplimentadas por estudiantes de grado y máster de la UCM en el curso 2016-2017. La muestra de profesorado incluye las respuestas de los 666 profesores participantes en el Programa DOCENTIA-UCM que impartían clase en grado y máster durante el curso 2016-2017.

Los análisis que se presentan están centrados en la dimensionalidad de los cuestionarios aplicados mediante Análisis Factorial Exploratorio (AFE), aportando una aproximación analítica actualizada que se alinea con las buenas prácticas recomendadas en el presente para los estudios que hacen uso de AFE de ítems ordinales en estudios de validación.

Los resultados muestran que la estructura interna del cuestionario para alumnado se define sobre la base de un solo factor que explica 73,4% de la varianza, denominado *Calidad de la Docencia*. Los resultados para el cuestionario del profesorado sugieren una configuración bidimensional del constructo *calidad de la docencia*, explicando el 63,912% de la varianza extraída. Estos dos factores se denominaron *Competencia docente general* y *Desarrollo específico de la propuesta docente*, que muestra una perspectiva más diferenciada o matizada que el alumnado en su visión del constructo.

PALABRAS CLAVE

Evaluación del profesorado; calidad de la docencia; educación superior; análisis de la dimensionalidad; Análisis Factorial Exploratorio.

ABSTRACT

After 75 years of research on student assessment questionnaires on teaching, the debate on their dimensionality is still open. This question has important implications, both theoretically (for the validation of the “quality of teaching” construct) and practically (with advantages associated with one-dimensional measures for summative purposes and multi-dimensional measures for training purposes).

The present research proposes an analysis of the dimensionality of a standard questionnaire for the evaluation of teaching by students; as well as completing the analysis of the quality of teaching construct by adding the perspective of the teaching staff, deepening the nature of what some perceive as quality teaching.

The double perspective adopted in this study is based on the use of the same teaching activity assessment questionnaire with the two groups involved, professors and students at the Complutense University of Madrid (UCM). A total of 125,155 surveys completed by UCM undergraduate and master students in the 2016-2017 academic year are analyzed. The sample of professors includes the answers of the 666 professors participating in the DOCENTIA-UCM Program who taught undergraduate and master's courses during the 2016-2017 academic year.

The analyses presented are focused on the dimensionality of the questionnaires applied by means of Exploratory Factor Analysis (AFE), providing an updated analytical approach that is aligned with the good practices recommended in the present for studies that make use of AFE of ordinal items in validation studies.

The results show that the internal structure of the student questionnaire is defined on the basis of a single factor that explains 73.4% of the variance, called Teaching Quality. The results for the teacher questionnaire suggest a two-dimensional configuration of the teaching quality construct, explaining 63.912% of the variance extracted. These two factors were called General teaching competence and Specific development of the teaching proposal, which shows a more differentiated or nuanced perspective than the students in their view of the construct.

KEY WORDS

Teacher evaluation; quality of teaching; questionnaires; factor structure; factor analysis; higher education.

INTRODUCCIÓN

Las primeras experiencias más o menos sistemáticas de la valoración de la docencia por el alumnado, mediante cuestionarios similares a los empleados desde hacía casi 50 años en el contexto anglosajón, se inician en España a finales de los años 80 del pasado siglo (Jornet, Suárez, González, y Pérez, 1996; Tejedor y García-Valcárcel, 1996). Desde entonces estas

prácticas se han ido extendiendo y generalizando, muy especialmente a partir del año 2007 cuando la Agencia Nacional de Evaluación de la Calidad y Acreditación (ANECA) pone en marcha el Programa DOCENTIA de Apoyo a la Evaluación de la Actividad Docente del Profesorado Universitario (Tejedor y Jornet, 2008). El objetivo de dicho programa es apoyar a las universidades en el diseño de mecanismos propios para gestionar la calidad de la actividad docente del profesorado universitario y favorecer su desarrollo y reconocimiento. En la actualidad más del 90% de las universidades españolas participan en DOCENTIA, a lo largo de sus distintas fases (ANECA, s.f).

De modo paralelo al desarrollo y generalización de la evaluación de la actividad docente, la comunidad científica ha ido prestando atención a este ámbito de la investigación evaluativa.

Un nutrido grupo de los trabajos publicados en nuestro país sobre evaluación de la docencia son de carácter teórico, principalmente centrados en la propuesta de modelos de evaluación, en la descripción o el análisis de enfoques evaluativos diversos (p.e. Escudero, 2000; Jornet, González, y Bakieva, 2012; Mayorga y Madrid, 2008; Murillo, 2008; Perales, Jornet, y González, 2014; Rodríguez-Gómez, 2000; Ruiz, 2005; Tejedor, 2003, 2009). Pero también se han ido desarrollando investigaciones empíricas específicamente centradas en las evaluaciones de los estudiantes mediante cuestionarios (véase, entre otros, Apodaca y Grad, 2002; Casero, 2008, 2016; Mallorga, Gallardo, y Madrid, 2016; López-Cámara, González-López, y de León-Huertas, 2015; Lizasoain, Etxeberria, y Lukas, 2017; Molero y Ruíz, 2005; Muñoz, Ríos de Deus, y Abalde, 2002; Ordóñez y Rodríguez, 2015; Pozo, Giménez, y Bretones, 2009). Esta última línea de trabajo se integra así en la amplísima tradición que en otros contextos tiene esta área de investigación (para una visión de conjunto véase Benton y Cashin, 2014; Onwuegbuzie, Daniel, & Collins, 2009; Spooren, Brockx, & Mortelmans, 2013; Spooren, Vandermoere, Vanderstraeten, & Pepermans, 2017).

Tras 75 años de investigación empírica sobre los cuestionarios de evaluación de los estudiantes sobre la enseñanza, de un modo sintético puede decirse que, tomados en conjunto, los resultados de la investigación disponible sugieren que estos tienden a ser estadísticamente fiables, razonablemente válidos en varios sentidos y relativamente libres de sesgos, quizá más que cualquier otro tipo de datos usados en la evaluación del profesorado. Adicionalmente también se cuenta con evidencia sobre su utilidad efectiva para mejorar la actividad docente, especialmente cuando se combinan con procesos de auto-reflexión y de asesoramiento o consulta. Pero lo anterior no significa que no haya cuestiones pendientes

de investigación, interrogantes abiertos y áreas en las que se necesita de investigación adicional (Alhija, 2017).

Dimensionalidad del constructo *calidad de la docencia* desde la perspectiva del alumnado

Un debate abierto al que se ha venido prestando especial atención es el de la dimensionalidad subyacente a los cuestionarios de evaluación de los estudiantes (Apodaca y Grad, 2005; Spooren, *et al.*, 2013). Se ha entendido que esta cuestión tiene importantes repercusiones, tanto teóricas (puesto que se centra en definitiva en la definición y validación del constructo “calidad de la docencia”) como prácticas (asociándose ventajas a las medidas unidimensionales para fines sumativos y a las multidimensionales para fines formativos).

En la revisión llevada a cabo por Spooren, *et al.* (2013) se sintetiza con claridad el estado actual de la investigación en este punto. En primer lugar, hay un acuerdo general en reconocer una estructura multidimensional de los cuestionarios de evaluación del alumnado sobre la docencia, que deberían orientarse a captar los múltiples aspectos que definen la buena práctica docente y que proporcionarían información útil para la evaluación formativa. Sin embargo, se carece de acuerdo en torno al número y naturaleza de tales dimensiones (que deberían tener sustento tanto teórico como empírico), lo que da como resultado que la investigación empírica varíe muy ampliamente en el número y la definición de las dimensiones propuestas. Los 11 trabajos revisados por estos autores, publicados entre 2000 y 2010 ilustran bien este hecho: el número de factores definidos osciló entre 2 y 12. Pero además, como igualmente señalan Spooren, *et al.* (2013), la investigación más reciente revela que muchas de las dimensiones subyacentes a los cuestionarios parecen estar afectadas por un constructo global (unidimensional).

La investigación en España sobre esta cuestión ha arrojado resultados similares en términos de variabilidad. En la tabla 1 puede verse una relación de estudios que presentan específicamente análisis de la dimensionalidad de los cuestionarios del alumnado sobre la enseñanza haciendo uso de Análisis Factorial Exploratorio, la aproximación metodológica mayoritaria a esta cuestión casi sin excepciones (Apodaca y Grad (2002) aplicaron Análisis Factorial Confirmatorio y Escalamiento Multidimensional).

Aunque los trabajos revisados en la tabla 1 presentan algunas diferencias obvias (p.e. número de ítems que conforman los cuestionarios, que varían entre 18 y 40) comparten el uso de instrumentos estándar de

evaluación notablemente similares entre sí (y de hecho, a los usados en todo el mundo), adoptando un formato de ítems con escala tipo-likert y 5 respuestas graduadas en todos los casos.

Tabla 1

Análisis de la dimensionalidad de los cuestionarios de evaluación para estudiantes sobre la docencia mediante Análisis Factorial Exploratorio. Estudios desarrollados en el contexto universitario español

Autor	Nº dimensiones (% de varianza total explicada)	Dimensiones (% de varianza explicada)
Muñoz, et al. (2002)*	7 (65,0%)	I. (42,8%) II. (5,9%) III. (4,1%) IV. (3,7%) V. (3,2%) VI. 2,8%) VII. (2,6%)
Molero y Ruíz (2005)	5 (59,9%)	I. Metodología docente (18,3%) II. Diseño de la Guía (13,2) III. Actitud del profesorado (11,3%) IV. Coherencia interna de los recursos didácticos (9,8%) V. Información sobre sistemas de evaluación (7,3%)
Casero (2008)**	6 (55,4%)	I. Conocimiento II. Evaluación III. Interrelación con la materia y Metodología IV. Programa V. Cumplimiento de obligaciones VI. Satisfacción
Pozo, et al. (2009)	2 (68,9%)	I. Docencia de calidad (45,6%) II. Compromiso docente (23,3%)
López-Cámara, et al. (2015)	4 (64,8%)	I. Interacción con el alumnado (50,4%) II. Metodología (5,7%) III. Obligaciones docentes y evaluación (4,5%) IV. Medios y recursos (4,2%)
Moreno-Murcia, et al. (2015)	3 (68,2%)	I. Planificación (38,0%) II. Desarrollo (15,4%) III. Resultados (14,8%)

Autor	Nº dimensiones (% de varianza total explicada)	Dimensiones (% de varianza explicada)
Casero (2016)	4 (64,7%)	I. Cualidades personales (25,5%) II. Competencia docente (24,0%) III. Cumplimiento formal (10,0%) IV. Materia (5,2%)
Lizasoain-Hernández, et al. (2017)***	1 (54,3%)	-

* Se anotan los valores correspondientes a la extracción, puesto que la solución obtenida no pudo ser rotada por problemas de convergencia.

**Se toman los resultados disponibles correspondientes a las formas estructuradas y en la primera solución factorial obtenida sobre el instrumento original (no disponible % de varianza para los 6 factores).

***A continuación se lleva a cabo un Análisis Factorial Confirmatorio sobre la misma muestra que presenta un ajuste satisfactorio para el modelo de tres factores hipotetizado.

Sin embargo, un aspecto común en la investigación desarrollada hasta la fecha podría ser problemático. Nos referimos a que desde el punto de vista analítico, siete de los ocho trabajos citados adoptan una aproximación idéntica, que podríamos calificar de “clásica”, al análisis de la dimensionalidad: realizan una Análisis de Componentes Principales (ACP) sobre la matriz de correlaciones de Pearson, tras el que se emplea el criterio de Kaiser para la retención de factores y se procede a una rotación ortogonal Varimax. Esta estrategia homogénea de análisis podría no estar arrojando resultados precisos sobre la dimensionalidad de los cuestionarios de evaluación.

Efectivamente esta práctica y algunas decisiones tradicionalmente asociadas a la misma se alejan de las recomendaciones actuales sobre el uso de Análisis Factorial Exploratorio de ítems en los estudios de validación (véase p.e. Bandalos y Finney, 2018; Izquierdo, Olea, y Abad, 2014), que invitan firmemente a evitar en particular el uso del *pack* denominado “*Little Jiffy*”: Análisis de Componentes Principales más Kaiser más Varimax (Lloret-Segura, Ferreres-Traver, Hernández-Baeza, y Tomás-Marco, 2014). Así, por ejemplo, se ha señalado que el uso de ACP aumenta espuriamente las cargas factoriales, los porcentajes de varianza explicados por los factores, y puede producir una sobreestimación de la dimensionalidad del conjunto de ítems. La literatura es igualmente unánime en aconsejar el uso de varios criterios objetivos para la retención de factores (p.e. test MAP, Análisis Paralelo, índices de ajuste tipo GFI o RSMR) y subraya que salvo justificación clara de independencia de factores, siempre se debería comenzar por evaluar soluciones rotadas oblicuamente.

Por tanto, parece necesario y oportuno el desarrollo de nuevos trabajos empíricos sobre la dimensionalidad de los cuestionarios de evaluación de los alumnos sobre la enseñanza que adopten las buenas prácticas recomendadas en los últimos años por la literatura especializada.

Dimensionalidad del constructo *calidad de la docencia* desde la perspectiva conjunta de estudiantes y profesorado

Dado el extenso uso de las evaluaciones del alumnado en todo el mundo, la investigación se ha centrado esencialmente en el análisis del constructo calidad de la docencia desde su perspectiva. Sin embargo ello no significa que tradicionalmente no haya habido una línea de tensión derivada de las posibles y plausibles diferencias entre lo que el profesorado y los estudiantes consideran una docencia de calidad en la universidad. Brown (1980, citado en Feldman, 1988) se refirió tempranamente a esta cuestión en términos poco ambiguos:

From the very beginning of their use, faculty have expressed reservations about the meaning (validity) of student responses regarding teaching effectiveness. Put simply, faculty have argued that they and students use different criteria in evaluating teaching. Naturally, faculty view their own standards as being more relevant for, or consistent with, the long-run mission of higher education (Brown, p. 234).

Este debate, por tanto, se cruza específicamente con la discusión sobre la dimensionalidad del constructo, pues la idea de qué componentes definen una buena enseñanza y cómo se estructuran puede depender de si los encuestados son estudiantes o docentes (Onwuegbuzie, *et al.*, 2007). Sin embargo, no son muchos los estudios empíricos que analizan conjuntamente las valoraciones de estudiantes y profesores sobre los elementos clave de una enseñanza de calidad.

Feldman (1988) revisó 31 estudios desarrollados en Estados Unidos y Canadá que incorporaban las percepciones conjuntas de alumnado y del profesorado sobre las características que definen una buena enseñanza. Sobre un total de 22 dimensiones analizadas, los resultados apuntaron que, en general, ambos colectivos presentan un alto grado de similitud en la importancia concedida a los atributos que definen una docencia de calidad. No obstante se hallaron tres notables diferencias. Por un lado, los estudiantes parecen conceder una mayor importancia que los profesores a la capacidad para despertar y promover el interés por parte de los docentes. Por otro, los alumnos conceden menos importancia que los profesores a la capacidad para promover el aprendizaje autónomo y también el desafío

intelectual y el pensamiento independiente. Por su parte Goldstein y Benassi (2006) aplicaron un cuestionario idéntico, centrado en la valoración de la calidad de dos tipos de docencia (magistral y centrada en el debate), a una muestra de alumnado y profesorado. Encontraron esencialmente la misma estructura factorial, lo que de acuerdo con los propios autores indica que estudiantes y profesores comparten creencias comunes sobre los atributos de los profesores excelentes.

En el contexto universitario español solo hemos identificado un trabajo que analice de forma conjunta las percepciones de profesorado y alumnado sobre un cuestionario común de evaluación de la calidad de la docencia. En el trabajo de Casero (2008) ambos colectivos valoraron la importancia de 92 ítems, tomados del trabajo de Muñoz, Ríos, y Abalde (2002), no hallando diferencias significativas en la importancia concedida al 70% de los mismos. Los profesores (N=148) otorgaron mayor importancia a la dimensión de cumplimiento de las obligaciones, mientras que el alumnado (N=1088) presentó valoraciones medias superiores en aspectos relativos a la evaluación, la satisfacción y las prácticas. En el análisis factorial exploratorio realizado para cada colectivo con los 32 ítems que reunieron mayor consenso (Análisis de Componentes Principales más Kaiser más Varimax), se hallaron soluciones multidimensionales en ambos casos pero con configuraciones sensiblemente diversas: 6 factores en la muestra de estudiantes y 8 en la de profesorado.

Dada la estrategia analítica empleada y el modesto tamaño de la muestra de profesorado para la realización de AFE, sería de alto interés la aportación de evidencia complementaria sobre la dimensionalidad del constructo *calidad de la docencia* percibida por el profesorado en la universidad española.

El presente estudio

En el contexto de los antecedentes descritos en los apartados anteriores, la presente investigación tiene un doble objetivo.

En primer lugar, se propone un análisis de la dimensionalidad de un cuestionario estándar de evaluación de la docencia por parte del alumnado, sumándose a una línea de trabajo notablemente desarrollada en nuestro país pero aportando una aproximación analítica actualizada que se alinea con las buenas prácticas recomendadas en el presente para los estudios que hacen uso de Análisis Factorial Exploratorio de ítems en estudios de validación. Se trata así de avanzar sobre la masa crítica existente relativa a la estructura y naturaleza del constructo *calidad de la docencia* percibida

por el alumnado desde parámetros renovados que puedan arrojar evidencia adicional sobre el problema.

En segundo lugar, la investigación se propone completar el análisis del constructo *calidad de la docencia* añadiendo la perspectiva del profesorado, agente esencial de los procesos objeto de evaluación, con el objetivo de determinar posibles diferencias o similitudes en la dimensionalidad y, por tanto, profundizar en la naturaleza de lo que unos y otros perciben como una docencia de calidad. Esta línea de trabajo cuenta con escasos antecedentes en España, por lo que este trabajo representa una aportación relevante y novedosa a la investigación sobre el tema.

La doble perspectiva adoptada en este estudio se sustenta en el uso del mismo cuestionario de valoración de la actividad docente con los dos colectivos implicados, profesores y estudiantes. Se emplean, concretamente, el cuestionario de alumnos de la Universidad Complutense de Madrid (UCM) y una forma paralela del mismo usada en los procesos de autoevaluación del profesorado, ambos empleados en el marco del programa DOCENTIA. De esta forma el estudio de la dimensionalidad en cada uno de estos grupos permite contar con una información estrictamente comparable.

METODOLOGÍA

Participantes

En el estudio se analizan un total de 125.155 encuestas cumplimentadas de forma voluntaria por estudiantes de grado y máster de la UCM en el curso 2016-2017 de todas las áreas de conocimiento y titulaciones impartidas. La muestra de profesorado incluye las respuestas de los 666 profesores participantes en el Programa DOCENTIA-UCM que impartían clase en grado y máster durante el curso 2016-2017. La participación en dicho programa se realizó por invitación y exigía haber participado previamente con resultados positivos en el Programa de Evaluación de la Actividad Docente en los dos últimos cursos. La cumplimentación del cuestionario fue obligatoria para dicho colectivo.

Instrumentos

El cuestionario del estudiante DOCENTIA-UCM consta de 17 preguntas valoradas en una escala de 0 (totalmente en desacuerdo) a 10 (totalmente de acuerdo), a la que se añade la posibilidad de contestar *no sabe*. El cuestionario de autoevaluación del profesorado incluye las mismas 17 preguntas que el

aplicado a los estudiantes, si bien estas guardan la preceptiva concordancia con el sujeto que las responde. No incluye la opción *no sabe*. Al profesorado se le solicitó que cumplimentara el cuestionario teniendo en cuenta, de modo global, el conjunto de su docencia en los tres últimos cursos.

Procedimiento

Los cuestionarios de evaluación de los estudiantes se administraron *online* durante los dos periodos de encuestas habilitados para el alumnado por la UCM (diciembre y mayo) en el año académico de referencia. Los cuestionarios del profesorado fueron contestados *online* por todos los participantes en el programa DOCENTIA-UCM a lo largo del mes de marzo de 2017.

Análisis de datos

De acuerdo con los objetivos principales de este estudio, los análisis que se presentan están centrados en la dimensionalidad de los cuestionarios aplicados a la muestra de alumnado y profesorado mediante Análisis Factorial Exploratorio. Como paso previo se llevaron a cabo análisis de diagnóstico de supuestos y de la distribución de las variables-ítems incluidas en los cuestionarios. Igualmente se lleva a cabo un estudio de la fiabilidad de las escalas.

Como primer paso se realizaron diagnósticos de normalidad multivariada de las distribuciones de los ítems y se inspeccionaron las matrices de correlaciones bivariadas para descartar patrones anómalos y no lineales. A continuación todos los análisis factoriales se realizaron siguiendo las siguientes pautas:

- Factorización de una matriz de correlaciones y de un método de extracción de factores ajustados a la naturaleza y distribución de los datos.
- Valoración de la adecuación de los datos para su factorización mediante la medida de adecuación muestral de Kaiser-Meyer-Olkin (KMO). La adecuación de los datos al AFE se considera “satisfactoria” cuando toma valores mayores a .80.
- Uso de varios criterios objetivos para la retención de factores, que combinados con criterios teóricos y de interpretabilidad conduzcan a la solución factorial más parsimoniosa y plausible. En nuestro

estudio particularmente se lleva a cabo la valoración del ajuste de soluciones alternativas mediante los índices GFI y RMSR. Se consideran indicativos de un ajuste satisfactorio valores de GFI iguales o superiores .95. Para RMSR valores iguales o inferiores a .05 son considerados satisfactorios y por debajo de .08 aceptables (Hu & Bentler, 1999).

- Uso de métodos de rotación oblicuos, empleando diversos procedimientos que permitan seleccionar como mejor criterio de rotación el que ofrezca la solución factorial más simple e informativa.
- Definición de factores que presentaran un mínimo de 3-4 variables con saturaciones superiores a .35 y una adecuada discrepancia de las saturaciones entre factores.

La estimación de la fiabilidad de las escalas y subescalas incluyó el cálculo de coeficientes ajustados a la métrica ordinal de las variables: alpha de Cronbach ordinal y theta de Armor ordinal (Elosúa y Zumbo, 2008; Gadermann, Guhn, & Zumbo, 2012).

Los análisis se realizaron con IBMSPSS 23 y los paquetes *psych*, *polycor*, *GPArotation*, *nFactors*, *corpcor*, *ICS* y *R.utils* de R 3.1. Se hizo uso del software *R essentials* para vincular R con SPSS y el complemento *RFactor* (Basto y Pereira, 2012).

RESULTADOS

Análisis preliminares y decisiones iniciales sobre la matriz y el método de factorización

El estudio inicial de los ítems de los cuestionarios, tanto del profesorado como del alumnado, reveló la falta de normalidad multivariada de las distribuciones. En el diagnóstico de normalidad multivariada para la muestra de alumnado se halló un valor de Asimetría igual a 877567,509 ($p = 0,000$) y Curtosis igual a 5176012,529 ($p = 0,000$). Para la muestra de profesorado los resultados fueron: Asimetría = 2007,587 ($p = 0,000$) y Curtosis igual a 5067,841 ($p = 0,000$). En consecuencia, se tomaron en ambos casos dos decisiones analíticas derivadas de esta situación inicial: a) la factorización de la matriz de correlaciones policóricas, y b) el empleo del procedimiento de Mínimos Cuadrados Generalizados (GLS) en el proceso de factorización.

De este modo se garantizaba un AFE ajustado a la naturaleza estricta de la métrica de las variables (ordinal) y a su distribución no normal, lo que desaconseja particularmente el uso del método de estimación de Máxima Verosimilitud (ML).

Análisis Factorial Exploratorio del cuestionario del alumnado

El índice KMO=0,971 alcanzó un valor excelente, lo que permitía continuar el proceso con garantías. Los resultados iniciales obtenidos en el AFE se muestran en la tabla 2.

Los resultados sugirieron que una solución unifactorial podría ser la más adecuada (tabla 2). Efectivamente los índices de ajuste del modelo de un factor son claramente satisfactorios: GFI= .992 y RMSR=.039. En todo caso se inspeccionó una solución de dos factores que, aunque presenta una muy ligera mejora en los índices de ajuste (GFI=.995 y RMSR=.025), incorpora un factor débil con un aporte menor de varianza explicada y sin una nítida interpretación sustantiva (lo que Lloret-Segura, *et al.*, 2014 denominan *factor común menor*). Por tanto, de acuerdo con las recomendaciones actuales, tras obtener una primera solución inicial satisfactoria e interpretable, se comparó con alternativas que podían parecer plausibles y se decidió la solución final atendiendo a criterios tanto objetivos como de significación teórica. El modelo unifactorial, con un ajuste satisfactorio, más parsimonioso y nítido sustantivamente, fue el seleccionado.

Tabla 2

Varianza explicada (autovalores iniciales) para el cuestionario del alumnado (extracto)

	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción		
	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado
1	13,323	78,373	78,373	13,155	77,382	77,382
2	,707	4,158	82,531			
3	,524	3,081	85,612			
4	,407	2,391	88,003			
5	,384	2,261	90,264			
6	,276	1,623	91,887			
17	,028	,162	100,000			

Por tanto, la estructura interna del cuestionario para alumnado se define sobre la base de un solo factor o dimensión que explica 73,4% de la varianza, un porcentaje sustancial, y que puede ser denominado *Calidad de la Docencia*. Dicho constructo se presenta como esencialmente unidimensional y unitario a la vista de estos resultados.

Los pesos o cargas factoriales de los ítems sobre el factor toman valores ciertamente satisfactorios, en el rango 0.75-0.95 (tabla 3).

El conjunto de comunalidades, en el rango .56-.91 indican igualmente que la totalidad de ítems también quedan satisfactoriamente representados en la solución propuesta (tabla 3).

Tabla 3

Matriz de patrones ordenada para el modelo unifactorial y comunalidades extraídas (cuestionario alumnado)

		Pesos factoriales	Comunalidades
Nota21	Estoy satisfecho con la labor docente del profesor	,952	,907
Nota20	En mi opinión es un buen profesor.	,947	,896
Nota19	La labor docente de este profesor me ayuda a adquirir conocimientos y competencias	,933	,870
Nota14	El profesor cumple con los horarios de clase establecidos	,913	,834
Nota8	El profesor explica de forma clara y comprensible	,911	,831
Nota1	El profesor informa de manera clara sobre los objetivos de la asignatura	,897	,805
Nota6	El profesor organiza y estructura bien las clases	,897	,804
Nota4	El profesor muestra competencia en la materia que explica	,895	,800
Nota18	El profesor despierta mi interés por la asignatura	,892	,796
Nota16	El profesor se muestra accesible con los estudiantes	,887	,787
Nota3	El profesor informa de manera clara sobre las actividades docentes	,862	,744

Nota17	El profesor mantiene un trato correcto con los estudiantes	,859	,738
Nota11	El sistema de evaluación permite al estudiante reflejar los conocimientos y competencias adquiridas	,858	,736
Nota2	El profesor informa de manera clara sobre el sistema de evaluación	,837	,701
Nota10	El profesor resuelve las dudas y ayuda a los estudiantes cuando lo necesitan	,824	,678
Nota9	Los materiales utilizados y/o recomendados son útiles para cursar la asignatura (bibliografía, material en Campus Virtual, etc.)	,815	,665
Nota12	El profesor cumple con el programa de la asignatura	,751	,563

Análisis Factorial Exploratorio del cuestionario para profesorado

El índice KMO=.940, muy elevado, confirmó la idoneidad de los datos para su factorización. Los resultados iniciales (tabla 4) aconsejaron inspeccionar y valorar soluciones de hasta tres factores.

Tabla 4

Varianza explicada (autovalores iniciales) para el cuestionario del profesorado (extracto)

	Total	% de varianza	% acumulado
1	10,229	60,168	60,168
2	1,209	7,113	67,281
3	,935	5,502	72,783
4	,863	5,077	77,860
5	,620	3,646	81,506
6	,495	2,914	84,420
17	,112	,661	100,000

Los índices de ajuste de las soluciones valoradas se muestran en la tabla 5.

Tabla 5
Índices de ajuste para tres modelos alternativos del cuestionario para profesorado

Modelo	GFI	RMSR
1 factor	.958	.071
2 factores	.971	.051
3 factores	.978	.034

El modelo de un factor presenta un ajuste aceptable, aunque con valores próximos al límite para considerarse satisfactorios. En los modelos de dos y tres factores se mejora sustancialmente el ajuste, calificable en ambos casos claramente de satisfactorio. Atendiendo a estas consideraciones se inspeccionaron en detalle y se valoraron soluciones de 2 y 3 factores haciendo uso de rotación oblicua por varios procedimientos (promax, geomin y quartimin) tal y como aconseja la literatura ya citada.

Un modelo de dos factores con rotación quartimin fue la solución más parsimoniosa e interpretable sustantivamente. De las soluciones con tres factores también la desarrollada con el método de rotación quartimin fue la más clara y consistente, pero presentó un factor común menor, definido de forma fuerte por solo dos ítems, y una estructura menos simple. Por tanto, se seleccionó como modelo más adecuado el bifactorial.

El modelo bifactorial propuesto explica el 63,912% de la varianza extraída (factor 1: 58,474% ; factor 2: 5,439%) y presenta una estructura razonablemente simple, como puede verse en la tabla 6. En la misma tabla pueden observarse las comunalidades extraídas que también presentaron valores globalmente satisfactorios, en el rango .45-.81, salvo para el ítem 11, con una comunalidad más modesta (.29).

La adscripción propuesta de los ítems a cada uno de los dos factores (destacados en negrita y sombreado) cumple con pocas excepciones los criterios más estrictos considerados en la literatura (véase Lloret-Segura, *et al.*, 2014). Estos incluyen considerar saturaciones iguales o superiores a .30 o .40; y considerar en los estudios empíricos que la discrepancia entre las saturaciones en los dos primeros factores sea del tipo .30/.50 o .60/.40, esto es, una diferencia de al menos .20. Como puede verse en la tabla 6 las cargas de todos los ítems que definen cada factor son iguales o superiores a .40, salvo en un caso (ítem 11), que es en todo caso superior a .30. En lo que respecta a la discrepancia entre las saturaciones, solo dos ítems presentan

una diferencia entre factores inferior a .20 (se trata de los ítems 10 y 11, con discrepancia de las saturaciones en todo caso $\geq .10$). La valoración teórico-conceptual, que se comenta ampliamente en el apartado de Discusión, ha sido usada en estos dos casos como criterio de adscripción complementario.

La correlación entre los dos factores definidos es de magnitud notable, igual a .76, lo que justifica la rotación oblicua en términos metodológicos.

Tabla 6

Matriz de patrones ordenada para el modelo bifactorial y comunalidades extraídas (cuestionario profesorado)

		Pesos factoriales		Comunalidades
		F1	F2	
Res20	Estoy satisfecho con mi labor docente del profesor	,953	-,073	,808
Res21	En mi opinión soy un buen profesor.	,913	-,095	,711
Res8	Explico de forma clara y comprensible	,829	,074	,785
Res18	Despierto el interés por las asignaturas que imparto	,809	-,024	,626
Res19	Mi labor docente ayuda a adquirir conocimientos y competencias a los estudiantes	,734	,182	,774
Res6	Organizo y estructuro bien las clases	,724	,189	,766
Res4	Soy competente en las materias que explico	,668	,143	,611
Res10	Resuelvo las dudas y ayudo a los estudiantes cuando lo necesitan	,439	,312	,498
Res2	Informo de manera clara sobre el sistema de evaluación	-,081	,920	,740
Res3	Informo de manera clara sobre las actividades docentes	-,036	,861	,695
Res14	Cumplo con los horarios de clase establecidos	,102	,744	,679

		Pesos factoriales		Comunalidades
		F1	F2	
Res12	Cumplo con el programa de la asignatura	-,104	,742	,445
Res16	Me muestro accesible con los estudiantes	,185	,631	,610
Res9	Los materiales que utilizo y/o recomiendo son útiles para cursar la asignatura (bibliografía, material en Campus Virtual, etc.)	,191	,592	,559
Res17	Mantengo un trato correcto con los estudiantes	,230	,588	,605
Res1	Informo de manera clara sobre los objetivos de la asignatura	,319	,543	,660
Res11	El sistema de evaluación de las asignaturas que imparto permite al estudiante reflejar los conocimientos y competencias adquiridas	,238	,339	,294

Fiabilidad de las escalas para el alumnado y el profesorado

La estimación de la fiabilidad para la escala completa correspondiente a ambas muestras, así como la correspondiente a las dos subescalas definidas para el profesorado fue altamente positiva (tabla 7). Se cuenta por tanto con escalas y subescalas de valoración de la calidad docente ciertamente fiables.

Tabla 7
Estimaciones de fiabilidad de escalas y subescalas

Escala	Coficiente alpha	Coficiente alpha ordinal	Coficiente theta (Armor)	Coficiente theta ordinal (Armor)
Alumnado	,974	,983	,977	,983
Profesorado	,929	,958	,932	,959
Factor I	,915	,944	,919	,945
Factor II	,854	,919	,863	,921

DISCUSIÓN

Los resultados de este estudio sugieren una configuración unidimensional del constructo *calidad de la docencia* en cuanto percibido por los estudiantes, a diferencia de lo encontrado en la mayor parte de la investigación previa en el contexto universitario español, que venía coincidiendo en su naturaleza multidimensional (Casero, 2008, 2016; López-Cámara, *et al.*, 2015; Molero y Ruiz, 2005; Moreno-Murcia, *et al.*, 2015; Muñoz, *et al.*, 2002; Pozo, *et al.*, 2009).

Para entender adecuadamente estos resultados es esencial considerar que todos los trabajos previos citados comparten una aproximación analítica en el AFE que en este estudio ha sido notablemente revisada para ajustarse a las buenas prácticas recomendadas actualmente (Bandalos & Finney, 2018; Izquierdo, Olea y Abad, 2014; Lloret, *et al.*, 2014). En primer lugar, se factorizó la matriz de correlaciones policóricas, con el fin de proporcionar un *input* ajustado a la métrica y distribución no normal de los ítems (muy frecuente en este tipo de instrumentos), frente al empleo de la matriz de correlaciones de Pearson (usada siempre en los estudios referenciados). En segundo lugar se descartó expresamente la aproximación Componentes Principales + Kaiser + Varimax, muy cuestionada en la actualidad y usada en todos los trabajos citados. Finalmente, en la interpretación de las soluciones factoriales se adoptó, junto con las consideraciones teórico-sustantivas oportunas, criterios técnicos estrictos para la definición de factores, con el fin de identificar dimensiones bien determinadas y estables. Puesto que el empleo de estrategias más o menos adecuadas en el AFE puede llevar a conclusiones notablemente diferentes (Lloret, Ferreres, Hernández, & Tomás, 2017), parece más que plausible que las discrepancias en nuestros hallazgos puedan estar derivadas en buena parte de la estrategia analítica empleada. Esta interpretación parece además reforzada por el hecho de que nuestros resultados sí son coincidentes con la estructura unidimensional hallada recientemente por Lizasoain-Hernández, *et al.* (2017) mediante AFE para el cuestionario de 18 ítems empleado en la Universidad del País Vasco. Estos autores también factorizaron la matriz de correlaciones policóricas y emplearon varios criterios objetivos para la retención de factores (coordenadas óptimas, factor de aceleración, análisis paralelo y el clásico criterio de Kaiser). Por otra parte, esta naturaleza en última instancia global (unidimensional) del constructo subyacente a los cuestionarios de evaluación de la docencia por parte de los estudiantes es coherente con la investigación y las concepciones teóricas recientes sobre el tema (Spooren, *et al.*, 2013). Y en este sentido es relevante subrayar que la proporción de varianza explicada por nuestra solución unifactorial (73,4%) es la más alta de las registradas en la literatura revisada, que varía entre 54,3% y 68,9% con soluciones de hasta 7 factores.

En todo caso los resultados obtenidos en este estudio deben ser considerados como un aporte más de evidencia en la agenda de investigación, que deben ser replicados y extendidos. En este sentido, es interesante notar que en el Análisis Factorial Confirmatorio posterior llevado a cabo por Lizasoain et al. (2017) se valoró también como plausible la estructura trifactorial hipotetizada, lo que abre una línea de indagación que consideramos de alto interés. Efectivamente aunque el ajuste informado del modelo fue satisfactorio, el tercer factor propuesto (Resultados) estuvo definido por dos únicos ítems y la correlación entre los dos factores restantes (Planificación y Desarrollo) fue de muy elevada magnitud (.92), lo que podría cuestionar su capacidad para representar de hecho constructos distintos. La valoración alternativa de un modelo unifactorial mediante AFC con nuestro cuestionario en una muestra distinta (tal y como se recomienda en la literatura) parece por tanto una línea de trabajo relevante derivada de los resultados obtenidos en este estudio. También permitiría profundizar en el estudio de la hipótesis de unidimensionalidad la valoración de modelos que prescindieran de los dos ítems de valoración genérica de la calidad docente (16 y 17), pues se ha sugerido que podrían contribuir a reforzar la misma (Apodaca y Grad, 2002).

Por otra parte, a diferencia de los hallazgos derivados de las valoraciones del alumnado, nuestros resultados sugieren una configuración bidimensional del constructo *calidad de la docencia* en cuanto percibido por el profesorado, cuya interpretación sustantiva es la que sigue:

- *Factor I. Competencia docente general.* Definido esencialmente por el juicio global que tiene el profesor de su labor docente, que se sustenta en sus percepciones sobre el dominio de la materia, la buena organización de la clase, la claridad expositiva, la adecuada resolución de dudas y el fomento efectivo del aprendizaje y el interés del alumnado.
- *Factor II. Desarrollo específico de la propuesta docente.* Definido esencialmente por las valoraciones sobre la adecuación del programa y la proporción de información sobre el mismo, el cumplimiento formal de tareas y la accesibilidad y trato con el alumnado.

El profesorado por tanto parece adoptar una perspectiva más diferenciada o matizada que el alumnado en su visión del constructo, lo que podría estar reforzado por el hecho de haber solicitado una auto-evaluación docente referida a los tres últimos cursos. En este escenario, el profesorado parece haber distinguido entre componentes que podríamos calificar de estructurales o “permanentes” (competencia docente general) y aquellos más funcionales o de ejecución (programa, cumplimiento, relacionales),

quizá más ligados a escenarios específicos (una asignatura, un curso, un grupo). En todo caso la notable correlación entre ambas dimensiones es coherente desde el punto de vista teórico-conceptual con una concepción de la calidad docente conformada por aspectos sustancialmente relacionados entre sí.

En conjunto nuestros hallazgos difieren notablemente de los informados por Casero (2008), único antecedente localizado en nuestro país, aunque conviene notar que ambos estudios son escasamente comparables y la limitación del tamaño de la muestra de este estudio anterior aconseja tomar los resultados del AFE con especial cautela. En todo caso, estos nuevos resultados sobre la calidad de la docencia percibida por el profesorado y su vinculación con las percepciones del alumnado sobre la base de un instrumento estándar común creemos que retoma una línea de trabajo interesante sin apenas antecedentes en nuestro país. Pero por ese mismo carácter inicial o de avance estos hallazgos requieren de réplica con otras muestras y quizá también con otros cuestionarios; también avanzar hacia estudios de dimensionalidad con vocación confirmatoria. En este sentido esperamos haber estimulado este campo de trabajo, que trata de contribuir a una mejor comprensión de las estructuras y procesos que subyacen a las prácticas evaluativas y, en última instancia, a que la evaluación de la actividad docente del profesorado se desarrolle en el marco de unas adecuadas garantías técnicas.

REFERENCIAS BIBLIOGRÁFICAS

- Alhija, F.N.A. (2017) Guest editor introduction to the special issue "Contemporary evaluation of teaching: Challenges and promises". *Studies in Educational Evaluation*, 54, 1-3.
- Apodaca, P., y Grad, H. (2002). Análisis dimensional de las opiniones de los alumnos universitarios sobre sus profesores: comparación entre técnicas paramétricas y no-paramétricas. *Revista de Investigación educativa*, 20(2), 385-409.
- Apodaca, P., y Grad, H. (2005). The dimensionality of student ratings of teaching: integration of uni-and multidimensional models. *Studies in Higher Education*, 30(6), 723-748.
- ANECA (s.f.). Docentia: descripción. <https://bit.ly/2H7xLv6>
- Bandalos, D.L., & Finney, S. J. (2018). Factor analysis: Exploratory and confirmatory. En G.R. Hancock & R.O. Mueller (Eds.), *The reviewer's guide to quantitative methods in the social sciences*. (pp. 93-104). Routledge.
- Basto, M., y Pereira, J. M. (2012). An SPSS R-menu for ordinal factor analysis. *Journal of statistical software*, 46(4), 1-29.
- Benton, S.L. y Cashin, W.E. (2014). Student ratings of instruction in college and university courses. En M.B. Paulsen (Ed.). *Higher Education: Handbook of Theory and Research*, v. 29, (pp. 279-326). Springer.
- Casero, A. (2008). Propuesta de un cuestionario de evaluación de la calidad docente universitaria consensuado entre alumnos y profesores. *Revista de Investigación Educativa*, 26(1), 25-44.
- Casero, A. (2016). Deconstrucción del "buen profesor". Análisis de las características formales y no formales de la docencia universitaria. *RELIEVE-Revista Electrónica de Investigación y Evaluación Educativa*, 22(2). <http://dx.doi.org/10.7203/relieve.22.2.9419>.
- Elosua, P. y Zumbo, B. (2008). Coeficientes de fiabilidad para escalas de respuesta categórica ordenada. *Psicothema*, 20(4), 896-901.
- Escudero, T. (2000). La evaluación y mejora de la enseñanza en la universidad: otra perspectiva. *Revista de Investigación Educativa*, 18(2), 405-416.
- Feldman, K.A. (1988). Effective college teaching from the students' and faculty's view: Matched or mismatched priorities? *Research in Higher Education*, 28(4), 291-329.
- Gadermann, A.M., Guhn, M., & Zumbo, B. (2012). Estimating ordinal reliability for Likert-type and ordinal item response data: a conceptual, empirical, and practical guide. *Practical Assessment, Research & Evaluation*, 17(3). <https://bit.ly/3bkRdmc>.
- Goldstein, G. S., & Benassi, V. A. (2006). Students' and instructors' beliefs about excellent lecturers and discussion leaders. *Research in Higher Education*, 47, 685-707.
- Hu, L., & Bentler, P. (1999). Cutoff criteria for fit indexes in Covariances Structures Analysis: conventional criteria versus new alternatives. *Structural Equation Modeling*, 61, 1-55.
- Izquierdo, I., Olea, J., & Abad, F.J. (2014). Exploratory Factor Analysis in validation studies: uses and recommendations. *Psicothema*, 26(3), 395-400. <http://dx.doi.org/10.7334/psicothema2013.349>

- Jornet, J., Suárez, J.M., González, J., y Pérez, A. (1996). Evaluación de la actividad universitaria. En G. Quintas (Coord.), *Educación, reforma y evaluación de la universidad*. (pp. 189-244). Universitat de Valencia.
- Jornet, J.; González Such, J., y Bakieva, M. (2012). Los resultados de aprendizaje como indicador para la evaluación de la calidad de la docencia universitaria. Reflexiones metodológicas. *Revista Iberoamericana de Evaluación Educativa*, 5(2), 100-115.
- Lizasoain-Hernández, L.; Etxeberria-Murgiondo, J., & Lukas-Mujika, J. F. (2017). A proposal for a new questionnaire for the evaluation of teachers at the University of the Basque Country. Dimensional, differential and psychometric study. *RELIEVE*, 23(2), art. 1. <https://doi.org/10.7203/relieve.23.2.10436>
- Lloret, S., Ferreres, A., Hernández, A., & Tomás, I. (2017). The exploratory factor analysis of items: Guided analysis based on empirical data and software. *Anales de Psicología*, 33(2), 417-432.
- Lloret-Segura, S., Ferreres-Traver, A., Hernández-Baeza, A. y Tomás-Marco, I. (2014). El análisis factorial exploratorio de los ítems: una guía práctica, revisada y actualizada. *Anales de Psicología*, 30(3), 1151-1169.
- López-Cámara, A. B., González-López, I., y de León-Huertas, C. (2015). Un análisis factorial exploratorio para la construcción de un modelo de indicadores de evaluación docente universitaria. *Cultura y Educación*, 27(2), 337-371.
- Mayorga, M.J. y Madrid, D. (2008). Evaluación de la docencia: aproximación a una realidad educativa en la universidad. *Enseñanza*, 26, 99-122.
- Mayorga, M.J.; Gallardo, M., y Madrid, D. (2018). Cómo construir un cuestionario para evaluar la docencia universitaria. Estudio empírico. *Universitas Tarroconensis. Revista de Ciències de l'Educació*, 2, 6-22. <https://doi.org/10.17345/ute.2016.2.974>
- Molero, D. y Ruíz, J. (2005). La evaluación de la docencia universitaria. Dimensiones y variables más relevantes. *Revista de Investigación Educativa*, 23(1), 57-84.
- Moreno-Murcia, J.A.; Silveira, Y., & Belando, N. (2015). Questionnaire evaluating teaching competencies in the university environment. Evaluation of teaching competencies in the university. *New Approaches In Educational Research*, 4(1) 54-61.
- Muñoz, J.M., Ríos de Deus, M.P. y Abalde, E. (2002). Evaluación docente vs. Evaluación de la calidad. *RELIEVE: Revista Electrónica de Investigación y Evaluación Educativa*, 8(2). <https://bit.ly/379s1M0>.
- Murillo, F.J. (2008). La Evaluación del Profesorado Universitario en España. *Revista Iberoamericana de Evaluación Educativa*, 1(3), 29-45.
- Onwuegbuzie, A. J., Daniel, L. G., & Collins, K. M. (2009). A meta-validation model for assessing the score-validity of student teaching evaluations. *Quality & Quantity*, 43(2), 197-209.
- Onwuegbuzie, A.J.; Witcher, W.A.E.; Collins, K.M.T.; Filer, J.D.; Wiedmaier, C.D., & Moore, C.W (2007). Students' perceptions of characteristics of effective college teachers: a validity study of a teaching evaluation form using a mixed-method analysis.

- American Educational Research Journal*, 44, 113-160.
- Ordóñez, R. y Rodríguez, M.R. (2015). Docencia en la universidad: valoraciones de los estudiantes de la Universidad de Sevilla. *Bordón. Revista de Pedagogía*, 3(67), 85-101.
- Perales, M.J., Jornet, J.M. y González, J. (2014). Tendencias en las políticas de formación y evaluación del profesorado en la Educación Superior en España. En *Revista Iberoamericana de Evaluación Educativa*, 2 (Especial), 53-64.
- Pozo, C., Giménez, M.L. y Bretones, B. (2009). La evaluación de la calidad docente en el nuevo marco del EEES. Un estudio sobre la encuesta de opinión del Programa DOCENTIA-ANDALUCÍA, *Educación XXI*, 11, 43-64.
- Rodríguez-Gómez, G. (2000). La evaluación de la actividad docente en la universidad: entre el sueño y la realidad. *Revista de investigación educativa*, 18(2), 417-432.
- Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching: The state of the art. *Review of Educational Research*, 83(4), 598-642.
- Spooren, P., Vandermoere, F., Vanderstraeten, & Pepermans, K. (2017). Exploring high impact scholarship in research on students evaluation of teaching (SET). *Educational Research Review*, 22, 129-141.
- Tejedor, F. J. y García- Valcárcel, A. (1996). La evaluación de la calidad de la docencia universitaria en el marco de la evaluación institucional desde la perspectiva del alumno. En F.J. Tejedor y J.L. Rodríguez Diéguez, (Coords), *Evaluación educativa II: Evaluación institucional fundamentos teóricos y aplicaciones prácticas*. (pp. 93-122). IUCE.
- Tejedor, F. J. y Jornet, J. M. (2008). La evaluación del profesorado universitario en España. *Revista Electrónica de Investigación Educativa, Especial*. <https://bit.ly/3besjVn>.
- Tejedor, F.J. (2003). Un modelo de evaluación del profesorado universitario. *Revista de Investigación Educativa*, 21(1), 157-182.
- Tejedor, F.J. (2009). Evaluación del profesorado universitario: enfoque metodológico y algunas aportaciones de la investigación. *Estudios sobre Educación- ESE*, 16, 74-102

PERFIL ACADÉMICO Y PROFESIONAL DE LOS AUTORES

María Castro Morera. ORCID: <http://orcid.org/0000-0002-2597-3621>

Catedrática de Métodos de Investigación y Diagnóstico en Educación. Líneas de investigación: evaluación de sistemas, programas y centros educativos. Directora del Grupo de Investigación “Medida y Evaluación de Sistemas Educativos” (M.E.S.E.). Email: macastro@ucm.es

Enrique Navarro Asencio. ORCID: <http://orcid.org/0000-0002-3052-146X>

Profesor Ayudante Doctor. Coordinador del Máster en Investigación en Educación. Miembro del grupo de investigación “Medida y Evaluación de Sistemas Educativos” (M.E.S.E.). Línea de investigación: medición educativa y evaluación del rendimiento académico y factores asociados. Email: enriquen@ucm.es

Ángeles Blanco Blanco †: <http://orcid.org/0000-0002-7556-9934>

Titular de Universidad. Línea de investigación: evaluación de sistemas educativos y validación y aplicaciones de la teoría cognitivo social del desarrollo de la carrera.

Fecha Recepción del Artículo: 04. Octubre. 2019

Fecha Aceptación del Artículo: 10. Noviembre. 2019

Fecha Revisión para Publicación: 27. Enero. 2020