

Propuesta de un nuevo cuestionario de evaluación de los profesores de la Universidad del País Vasco. Estudio psicométrico, dimensional y diferencial

A proposal for a new questionnaire for the evaluation of teachers at the University of the Basque Country. Dimensional, differential and psychometric study

Lizasoain-Hernández, Luis; Etxeberria-Murgiondo, Juan & Lukas-Mujika, José F.
Universidad del País Vasco (España)

Abstract

The aim of this paper is to analyze the new questionnaire designed by the University of the Basque Country (UPV/EHU) to evaluate its teaching staff (SET). To do it, the responses of a 941 students sample were analyzed and the following aspects of the questionnaire were studied: its reliability, dimensionality, construct and criterion validity; concluding with a differential study considering variables such as gender, disciplinary field, perceived difficulty level or subject interest. The results suggest high internal consistency that fits to the theoretical dimensions: planning, process and results, enabling a formative use of information.

Keywords:

Students' Evaluation of Teaching, SET; Dimensionality; Questionnaire; Student Evaluation of Teacher Performance; Teacher Competencies

Reception Date
2017 June 25

Approval Date
2017 October 25

Publication Date:
2017 October 26

Resumen

El objetivo de este artículo es analizar el proyecto de nuevo cuestionario diseñado por la Universidad del País Vasco (UPV/EHU) para realizar la evaluación de sus docentes (SET). Se analizan las respuestas de una muestra de 941 estudiantes y se estudia la fiabilidad del cuestionario, la dimensionalidad, la validez de constructo y criterial, finalizando con un estudio diferencial tomando en cuenta variables como el género, el campo disciplinar, el nivel percibido de dificultad o el interés de las materias. Los resultados permiten afirmar que se trata de un instrumento de alta consistencia interna que se ajusta a las dimensiones teóricas usadas para su diseño y construcción: planificación, proceso y resultados, lo que posibilita un uso formativo de la información

Palabras clave:

Evaluación de la actividad docente por estudiantes; Dimensionalidad; Cuestionario; Evaluación del desempeño docente; Competencias docentes

Fecha de recepción
2017 Junio 25

Fecha de aprobación
2017 Octubre 25

Fecha de publicación
2017 Octubre 26

La evaluación de la actividad docente por parte de los estudiantes (SET) en instituciones de educación superior tiene una larga trayectoria y autores como Otani, Kim y Cho (2012) datan sus orígenes a principios del siglo XX. Como en otras muchas universidades, la evaluación de la actividad docente de los profesores de la Universidad del País Vasco (UPV/EHU) se realiza mediante la aplicación de un cuestionario. Este tipo de evaluación se

inició con carácter experimental y voluntario en el curso 1988-1989. Desde entonces el procedimiento siempre ha sido el de la aplicación de un cuestionario mediante el cual los estudiantes evalúan la actividad desempeñada por los docentes.

Lógicamente, a lo largo del tiempo este instrumento ha ido experimentando diferentes modificaciones, pero desde hace ya bastantes

Autor de contacto / Corresponding author

Lizasoain-Hernández, Luis. Departamento de Métodos de Investigación y Diagnóstico en Educación.
Universidad del País Vasco (España). luis.lizasoain@ehu.es

años podemos decir que este procedimiento evaluativo se aplica en todos los centros y a todos los profesores. El mismo se ha incorporado con total normalidad a la actividad cotidiana de la universidad. No obstante, con la implantación de la metodología requerida por el Espacio Europeo de Educación Superior (EEES) el cuestionario utilizado fue modificado en el curso 2006/2007.

Más adelante, en 2010, la Junta de Gobierno de la UPV/EHU aprobó el modelo IKD que supone un proceso de enseñanza-aprendizaje cooperativo y dinámico centrado en el alumnado y que se configura como una propuesta para el desarrollo curricular de las enseñanzas. Este modelo se caracteriza por ser dinámico y activo, centrado en el aprendizaje del alumnado, plural y flexible, para que pueda ser adaptado a cada titulación y centro docente y cooperativo entre los diferentes agentes implicados en la docencia.

Bajo las premisas señaladas, se pretendió construir un cuestionario de opinión al alumnado sobre la docencia de su profesorado que estuviera adaptado al EEES y a DOCENTIA que es una propuesta de la ANECA (Agencia Nacional de Evaluación de la Calidad y Acreditación Española). El cuestionario se diseñó desde una perspectiva alineada con DOCENTIA, cuyo objeto es apoyar a las universidades en el diseño de mecanismos propios para gestionar la calidad de la actividad docente del profesorado universitario.

La construcción del nuevo instrumento se llevó a cabo siguiendo distintas estrategias, entre las que se pueden señalar la revisión bibliográfica de las competencias que debe tener el profesorado en el EEES y el análisis del modelo educativo de la UPV/EHU para tratar de determinar el modelo de profesorado al que se aspira (Lukas, Santiago, Etxeberria y Lizasoain, 2014).

Tras la revisión, se elaboró un documento marco en el que se distinguieron las tres dimensiones que había de tener el nuevo cuestionario, a saber, Planificación de la

docencia, Desarrollo de la enseñanza y Resultados. Las mismas coinciden con la propuesta de la ANECA que son contempladas en el programa DOCENTIA de calidad de la docencia universitaria. De la misma manera, se tuvieron en cuenta los tres momentos que se distinguen en el proceso de enseñanza-aprendizaje, es decir, la preparación del proceso por parte del docente, el desarrollo de la docencia y los resultados obtenidos como consecuencia de dicho proceso. Además de la definición de las dimensiones, en el documento se especificaron los indicadores del proceso de enseñanza-aprendizaje susceptibles de ser evaluados en cada una. Solo se incluyeron aquellos que pueden ser directamente apreciados por el alumnado. Finalmente, se presentaban los ítems clasificados por dimensiones. Este documento fue debatido y analizado en diferentes grupos de discusión con objeto de recoger las opiniones de diferentes agentes universitarios (expertos externos e internos, técnicos en evaluación, profesorado y alumnado) (Lukas, Santiago, Etxeberria y Lizasoain, 2014).

Tras los análisis de las transcripciones de los distintos grupos de discusión, el nuevo cuestionario se componía de 18 elementos (en una escala Likert de 5 valores) mediante los que se valoran las diferentes dimensiones mencionadas. Junto a estos, hay un último elemento utilizado como criterio de valoración global del docente (“En general, pienso que es un buen profesor o profesora”). A su vez se incluyen elementos de información contextual y de autoevaluación de los propios estudiantes. Entre estos, merece destacar los ítems que miden el nivel percibido de dificultad de la asignatura y el interés inicial y final por la misma. Estas percepciones han sido utilizadas después para el análisis comparativo de la calidad docente.

Los 18 ítems (ver figura 1) que conforman el cuestionario se estructuran alrededor de tres constructos teóricos relativos a otros tantos aspectos de la actividad docente: planificación (ítems 1-5), proceso y desarrollo de la actividad docente (6-16) y resultados (17-18).

El profesor o profesora:
<ol style="list-style-type: none">1. Proporciona una Guía del Estudiante (<i>programación</i>) útil para el seguimiento de la asignatura2. Planifica de manera ordenada la asignatura3. Recomienda recursos (<i>bibliografía, materiales, TICS, etc.</i>) que facilitan nuestro aprendizaje4. Comunica claramente al comienzo de curso los criterios y procedimientos de evaluación5. Adecúa el nivel de dificultad de las clases a nuestros conocimientos previos6. Explica de manera clara y ordenada7. Favorece la actitud reflexiva y crítica8. Propone actividades para favorecer el aprendizaje autónomo (<i>búsqueda de información, resolución de casos y problemas prácticos, investigaciones, etc.</i>)9. Estimula la colaboración y el trabajo en equipo10. Evalúa las actividades que realizamos y nos informa con el fin de que podamos mejorar11. Motiva al alumnado para que se interese por su proceso de aprendizaje12. Atiende las consultas que le planteamos13. Transmite entusiasmo impartiendo esta asignatura14. Estimula el desarrollo de la expresión oral y/o escrita15. Se esfuerza para que entendamos la conexión de la asignatura con el resto de asignaturas16. Utiliza una metodología y unos recursos didácticos que favorecen el aprendizaje..17. He aprendido mucho cursando esta asignatura18. He mejorado mi nivel de partida con relación a las competencias previstas <p>Ítem CRITERIO: En general, pienso que es un buen profesor o profesora</p>

Figura 1 - Ítems del cuestionario

En concreto, este trabajo tiene como objetivos el estudiar las propiedades psicométricas del nuevo instrumento y realizar un estudio dimensional y diferencial de las respuestas de los estudiantes. Específicamente se estudia la fiabilidad del cuestionario (entendida como consistencia interna de las respuestas), la dimensionalidad mediante análisis factorial exploratorio, la validez de constructo -en la que se ha incluido el análisis factorial confirmatorio- y la validez criterial. Por último, se realiza un estudio diferencial tomando en cuenta variables como el género y el nivel percibido de dificultad e interés de las materias.

Como antecedentes de este tipo de estudio se pueden citar los siguientes: en primer lugar, el trabajo de Muñoz, Ríos y Abalde (2002) en el que se realiza una revisión de los instrumentos empleados por las distintas universidades españolas. En segundo lugar, los trabajos de Apodaca y Grad (2002, 2005) se centran precisamente en el estudio de la dimensionalidad de una versión anterior del cuestionario empleado por la UPV/EHU y en los mismos se analizan las posibilidades y

limitaciones de las funciones sumativas y formativas de este tipo de instrumentos en relación a su estructura uni o multidimensional.

González (2003) estudió mediante técnicas factoriales las dimensiones subyacentes al constructo de calidad universitaria. Un estudio de las propiedades psicométricas empleando análisis factorial confirmatorio para evaluar el grado de ajuste a la estructura dimensional teóricamente establecida es el llevado a cabo por Ginns, Prosser y Barrie (2007) con datos del cuestionario empleado en la Universidad de Sidney. Igualmente, el trabajo de Lemos, Queirós, Teixeira y Menezes (2011) desarrolla y valida un cuestionario multidimensional de aplicación en la Universidad de Porto.

Casero (2010) analiza los factores moduladores de la percepción que los estudiantes tienen de la calidad de sus docentes y su relación con la validez del sistema de evaluación. Mortelmans y Spooren (2009) usan el análisis factorial confirmatorio para analizar la validez de constructo del

instrumento SET37 concluyendo que el ajuste a un modelo teórico multidimensional es compatible con un factor general subyacente al que denominan “profesionalidad docente”. Planteamiento y conclusiones similares son las de Burdsal y Harrison (2008) que estudian la validez de un perfil multidimensional y de una evaluación global de la eficacia docente. Enfoque parecido asumen Alvarado, Morales y Aguayo (2016) a la hora de analizar la percepción de la calidad educativa de los estudiantes de la Universidad Autónoma de Nuevo León y del Instituto Tecnológico de Monterrey.

Por otra parte, aspectos parciales de las propiedades métricas de este tipo de instrumentos, como la consistencia interna o temporal, son analizados por Zhao y Gallant (2012), Roberts (2011) y Chen y Watkins (2010).

La cuestión de la validez de este tipo de instrumentos es una preocupación recurrente en la literatura. Una exhaustiva revisión del estado del arte sobre SET y la validez se encuentra en el trabajo de Spooren, Brockx y Mortelmans (2013) basado en el modelo de meta-validación en el que revisan los informes de investigaciones publicados desde 2000 en revistas evaluadas por pares.

Junto a esta revisión, hay un conjunto de trabajos que abordan la cuestión crucial de la relación entre la satisfacción del estudiante y el nivel de aprendizaje logrado. El de Molero (2007) se centra en experiencias de implantación del EEES. Fernández, Esteban, Fernández, Álvarez y Martínez (2007) analizan los datos de la Universidad de Oviedo. En esta misma línea el estudio de Stark-Wroblewski, Ahlering y Brill (2007) concluye que las medidas proporcionadas por instrumentos de tipo SET y las correspondientes al aprendizaje logrado por los estudiantes, evalúan y se refieren a aspectos distintos de la actividad docente. Zerihun, Beishuizen y Van Os (2012) desarrollan un cuestionario basado en las experiencias de aprendizaje de los estudiantes. Finalmente, con respecto a esta cuestión, es

preciso citar el meta-análisis realizado por Uttl, White y Gonzalez (2016) en el que se concluye que no existe relación entre el nivel de aprendizaje de los estudiantes y las valoraciones que los mismos expresan o realizan relativas a sus docentes.

Por último, en lo relativo a los estudios de enfoque diferencial, Caldera, Carranza, Jiménez y Pérez (2015), en el contexto del diseño de una escala de medición de las actitudes de los estudiantes universitarios ante la tutoría, encuentran diferencias por sexo y tipo de carrera. Haarala-Muhonen, Ruohonieni, Katajavuori y Lindblom-Ylänne (2011) analizan las diferencias en las percepciones de los estudiantes de diferentes campos disciplinares en la Universidad de Helsinki sobre sus entornos de enseñanza-aprendizaje incluida la valoración de la calidad docente. Concluyen que los estudiantes de farmacia y veterinaria tendían a realizar valoraciones más positivas que los de derecho. Kember y Leung (2011) comparan mediante técnicas SEM las respuestas de muestras de estudiantes de la Universidad de Hong Kong concluyendo que se da un modelo común de lo que los estudiantes consideran buena enseñanza, pero con diferencias entre los cuatro grupos disciplinares considerados. El trabajo de Tomkiewicz y Bass (2008) analiza las diferentes percepciones entre estudiantes masculinos y femeninos con respecto a docentes de su mismo o diferente género. Pascual (2007) analiza mediante modelos jerárquicos la satisfacción de los estudiantes tomando en consideración el agrupamiento de los mismos en facultades y escuelas.

De Juanas y Beltrán (2014) analizan las diferencias en las percepciones de 1388 estudiantes de la Universidad Complutense de Madrid, en función del género y las calificaciones. Rantanen (2013) utiliza un enfoque multinivel para analizar la estabilidad de las valoraciones de los estudiantes y un análisis de modelos jerárquicos que revela diferencias significativas entre los distintos tipos de estudiantes al evaluar diferentes

estratos de docentes. Por último, Addison, Best y Warrington (2006) concretan la relación entre las percepciones de la dificultad de la asignatura por parte de los estudiantes y las valoraciones que los mismos realizan de sus docentes.

Metodología

Participantes

Este trabajo analiza las respuestas dadas por una muestra de estudiantes de la UPV/EHU que fue seleccionada considerando el estrato del campo científico (los 5 habituales) y el curso (1º y 3º). La unidad muestral fue el aula. La selección de las aulas se concretó teniendo en cuenta la distribución del alumnado en los distintos campos de conocimiento y niveles de curso. En total fueron 34 las aulas participantes y 998 los estudiantes. Una vez eliminados 4 casos con errores, en los análisis exploratorios iniciales se detectó la existencia de estudiantes que respondían a todos los ítems de manera uniforme. En total eran 53 casos (estudiantes) que en términos porcentuales representan el 5,3% del total. El procedimiento empleado para su detección ha consistido en seleccionar aquellos casos en los que la desviación típica de las respuestas fuese cero. La puntuación que mayoritariamente asignan es de 4 ó 5.

Una vez identificados, y habida cuenta de los objetivos de este estudio, estos casos fueron eliminados del archivo de datos y los análisis y resultados que a continuación se presentan se basan en la muestra final de 941 estudiantes cuyas respuestas tienen varianza, por pequeña que ésta sea. Es cierto que esta carencia de variabilidad en las respuestas se puede interpretar de muchas maneras, pero

para los objetivos de este trabajo parece más oportuno no tomar en cuenta a quienes valoran de manera constante a sus docentes en todos los ítems y aspectos. A este respecto puede consultarse el trabajo de Choi y Kim (2014) en el que se analizan los patrones de respuesta monotónica en universidades coreanas. Con una perspectiva distinta, Darby (2007) estudia el posible efecto halo en las respuestas y Adams y Umbach (2012), empleando procedimientos multinivel, analizan los posibles factores explicativos de la ausencia de respuesta en las evaluaciones online.

Los 941 estudiantes se distribuyen por campo científico tal y como muestra la tabla 1.

Tabla 1: Distribución de la muestra por campos científicos

Campo científico	Frecuencia	Porcentaje
Humanidades	187	19,9
Ciencias Experimentales	169	18,0
Enseñanzas Técnicas	182	19,3
Ciencias de la Salud	283	30,1
Ciencias Sociales y Jurídicas	120	12,8
Total	941	100,0

Por último, y con objeto de caracterizar más completamente a los participantes, la tabla 2 muestra la distribución de algunas variables sociodemográficas relevantes que se han incorporado en otros apartados de este estudio.

Tabla 2 – Características contextuales de los participantes

1. Curso del alumno/a							
1er curso	2º curso	3º curso	4º curso	5º curso	Otros		
69,5%	2,3%	28,2%					
2. Sexo del alumno/a							
Mujer				Hombre			
63,2% (61,7%)				36,8% (38,3%)			
3. Edad del alumno/a							
18	19	20	21	22	De 23 a 25	De 26 a 35	36 o más
25,0%	22,3%	15,6%	10,3%	4,0%	11,1%	8,7%	3,1%
4. Convocatoria en la asignatura							
1ª Conv.	2ª Conv.	3ª Conv.	4ª Conv.	5ª Conv.	6ª Conv.		
97,2% (93,8%)	1,6% (3,7%)	0,5% (1,7%)	0,3% (0,4%)	0,1% (0,3%)	0,2% (0,1%)		
5. Horas de estudio							
De 0 a 1		De 2 a 3		De 4 a 5		De 6 a 7	8 o más
31,9% (37,4%)		47,5% (44,3%)		15,7% (12,6%)		3,5% (3,1%)	1,4% (2,5%)
6. Nivel de dificultad de la asignatura							
Muy Difícil		Difícil		Normal		Fácil	Muy Fácil
5,0% (8,6%)		29,9% (33,5%)		56,4% (50,1%)		7,8% (6,8%)	0,9% (0,9%)
7. Interés inicial en la asignatura							
Muy Bajo		Bajo		Medio		Alto	Muy Alto
3,3% (3,5%)		13,6% (12,7%)		43,9% (44,3%)		32,3% (31,1%)	6,9% (8,3%)
8. Interés final en la asignatura							
Muy Bajo		Bajo		Medio		Alto	Muy Alto
4,0% (4,5%)		11,2% (10,1%)		39,6% (39,1%)		35,2% (35,9%)	9,9% (10,4%)

(entre paréntesis datos de la población)

Con respecto a la información resumida en estas dos tablas conviene señalar que la UPV/EHU es la única universidad pública en la Comunidad Autónoma del País Vasco y que se trata de una universidad generalista no especializada que imparte titulaciones en todos los campos científicos.

Procedimiento

El cálculo del parámetro de la fiabilidad como consistencia interna del cuestionario se ha llevado a cabo mediante el estadístico habitual (Alfa de Cronbach).

Con objeto de analizar la dimensionalidad del cuestionario, se procedió a efectuar un análisis de Componentes Principales (ACP). Dada la métrica de los ítems se ha realizado el análisis de componentes para variables ordinales empleando el paquete R-Factor (Basto y Pereira, 2012) y usando la matriz de correlaciones policóricas entre los ítems.

Con respecto a la validez de constructo, se ha seguido la estrategia analítica habitual realizando un Análisis Factorial Confirmatorio (AFC) mediante modelos de ecuaciones estructurales (SEM) lo que posibilita confirmar el grado de ajuste de las dimensiones definidas teóricamente.

El estudio de la validez del cuestionario desde la perspectiva del criterio se ha llevado a cabo analizando en primer lugar las correlaciones de las respuestas a cada uno de los ítems con las dadas al ítem criterio (“En general pienso que es un buen profesor o profesora”). A continuación, para analizar el orden y peso de cada ítem en la explicación del ítem criterio, se aplicó el método de regresión por pasos segmentando la muestra en función de los cinco grandes campos científicos.

Para finalizar, se lleva a cabo un estudio diferencial tomando en cuenta algunas variables incluidas en el cuestionario, tales como el género de los estudiantes, el nivel de

dificultad de las asignaturas y el interés inicial y final por las mismas. Para ello se estudian las diferencias en la tendencia central y variabilidad de las respuestas a los ítems del cuestionario en función de los grupos generados por estas variables.

Para todo ello se ha empleado el programa estadístico SPSS en su versión 23 con el subprograma AMOS para el AFC y con el módulo R-Factor incorporado.

Resultados

Análisis de la Fiabilidad del Cuestionario.

Como se ha señalado, el cálculo del parámetro de la fiabilidad se llevó a cabo mediante el Alfa de Cronbach y se calculó segmentando la muestra objeto de estudio en función del campo científico. Los resultados son similares en los diferentes campos científicos tal y como muestra la tabla 3.

Tabla 3 - Coeficiente de fiabilidad del cuestionario en las diferentes áreas

CAMPO CIENTÍFICO	Alfa de Cronbach	N de elementos
Ciencias Experimentales	0,937	18
Enseñanza Técnicas	0,918	18
Ciencias de la Salud	0,949	18
Ciencias Sociales y Jurídicas	0,939	18
Humanidades	0,944	18
TOTAL PARTICIPANTES	0,939	18

La correlación de cada elemento con el total es elevada, con una correlación media para los 18 de 0,657 siendo el ítem 1 el de correlación más baja (0,513) y el 11 el de más alta (0,766). No hay ningún elemento que, en caso de ser eliminado, su ausencia lleve aparejado un

incremento del valor del Alfa de Cronbach. De cualquier forma, la mayoría de las correlaciones se sitúan muy cerca del promedio (ver tabla 4). Todos estos índices evidencian que se trata de un instrumento de alta consistencia interna.

Tabla 4 - Estadísticos ítems-total del cuestionario

	Correlación elemento-total corregida	Alfa de Cronbach si se elimina el elemento
ítem1	,513	,938
ítem2	,696	,934
ítem3	,659	,935
ítem4	,631	,936
ítem5	,666	,935
ítem6	,724	,934
ítem7	,647	,935
ítem8	,620	,936
ítem9	,533	,938
ítem10	,578	,937
ítem11	,766	,933
ítem12	,655	,935
ítem13	,674	,935
ítem14	,674	,935
ítem15	,651	,935
ítem16	,744	,934
ítem17	,713	,934
ítem18	,679	,935

Dimensionalidad del cuestionario. Análisis factorial exploratorio.

Con respecto a la dimensionalidad del cuestionario, un primer indicador de la estructura dimensional de una prueba es el porcentaje de varianza que explica cada uno de los componentes generados por la solución

factorial. Como se observa en la Tabla 5, el primer componente explica, por sí solo, el 54,29% de la varianza total de la prueba. Otro indicador utilizado es el de los autovalores de los componentes, y en este caso, el segundo es igual a 1 y los siguientes son menores.

Tabla 5 - Porcentaje de varianza que explica cada uno de los componentes generados

	Desv. Tip.	Valores propios	% de la varianza	% Acum.
Comp.1	3,126	9,772	54,288	54,288
Comp.2	1,000	1,000	5,553	59,841
Comp.3	,976	,952	5,291	65,132
Comp.4	,879	,773	4,295	69,427

El programa R-Factor, además de calcular el ACP para variables ordinales, proporciona diversos criterios para establecer el número de componentes a retener: coordenadas óptimas, factor de aceleración, análisis paralelo y el habitual criterio de Kaiser: Cattell, 1966; Glorfeld, 1995; Ledesma y Valero-Mora, 2007; Peres-Neto, Jackson y Somers, 2005; Revelle y Rocklin, 1979; Velicer, Wayne,

Eaton y Fava, 2000; Zwick, William, Velicer y Wayne, 1986. La tabla 6 muestra como, para los 4 criterios, el número de componentes a retener es de 1.

La representación gráfica de la aplicación de los 4 criterios se encuentra representada en la figura 2.

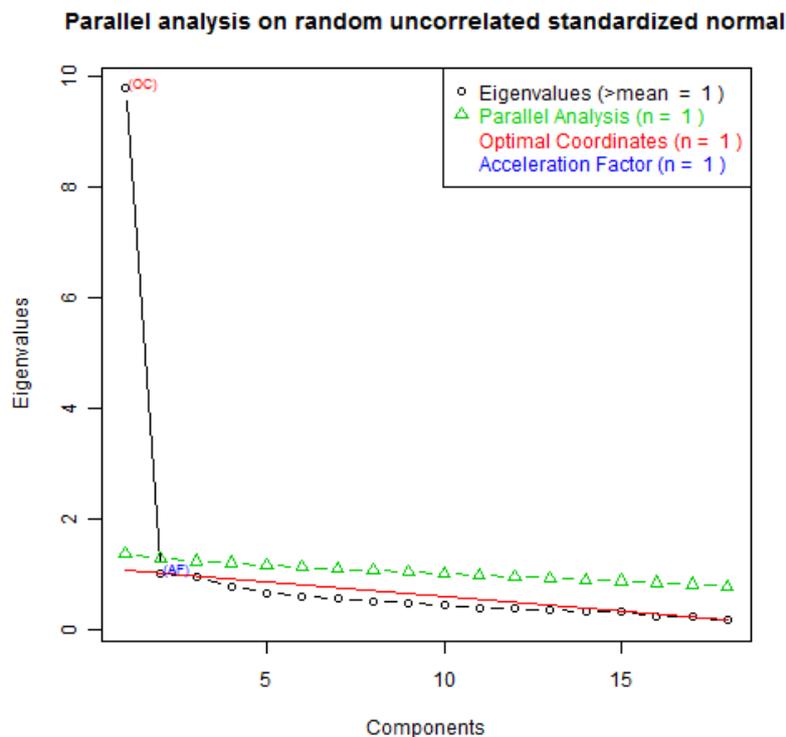


Figura 2 - Criterios de dimensionalidad.

Como conclusión de este apartado, el análisis de los cuestionarios cumplimentados por la muestra de estudiantes de la UPV-EHU, permite afirmar que se dispone de una fuerte evidencia que apunta a que el cuestionario que se emplea para valorar la práctica de los docentes es un instrumento de muy alta consistencia interna, con una adecuada capacidad discriminativa.

Validez de constructo

Como se ha mencionado previamente, a la hora de diseñar el cuestionario se tuvieron en cuenta diferentes constructos o dimensiones

teóricas que derivaron en indicadores que, a su vez, exigieron la construcción de ítems. Estas dimensiones son las siguientes: Planificación (ítems 1-5), Proceso y desarrollo de la actividad docente (6-16) y Resultados (17-18).

En este caso se ha realizado un Análisis Factorial Confirmatorio (AFC) mediante modelos de ecuaciones estructurales (SEM) lo que ha permitido evaluar el grado de ajuste de las dimensiones definidas teóricamente. En la figura 3 se muestra la estructura hipotetizada relativa a las tres dimensiones teóricas y los estadísticos asociados.

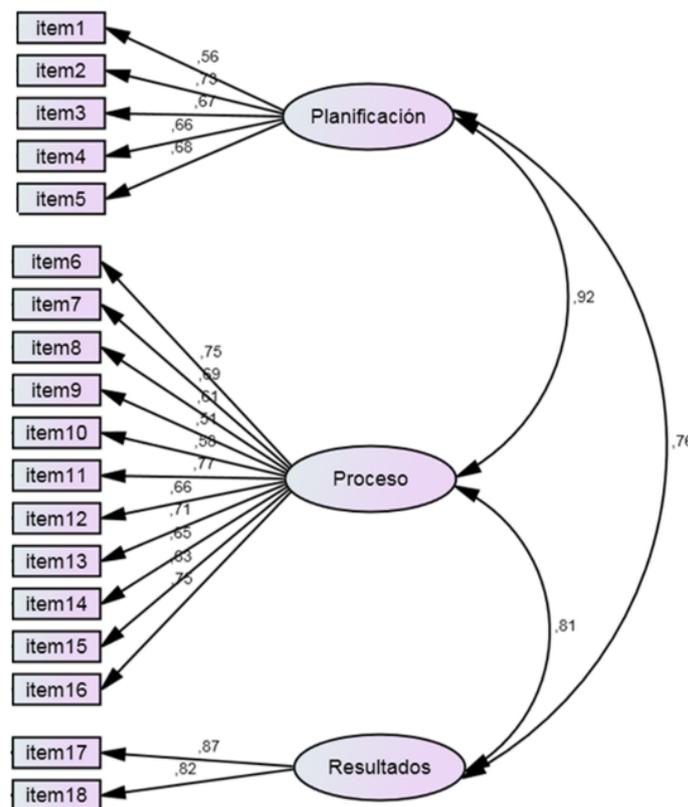


Figura 3 - Cargas factoriales y correlaciones entre los factores: Planificación, Proceso y Resultados

En la tabla 6 se muestran los estadísticos de ajuste empíricamente obtenidos y el valor límite habitual de los criterios de ajuste (Schreider et al., 2006; Hooper et al., 2008; Kaplan, 2009; Kline, 2011). En la misma se puede comprobar la plausibilidad del

modelo teóricamente establecido de forma que las respuestas de los estudiantes a los ítems del cuestionario se ajustan razonablemente a las dimensiones en que dichos ítems se agrupan teóricamente.

Tabla 6 - Estadísticos de ajuste del AFC

	Valores obtenidos	Valores criterio habituales
CFI	0,955	> 0,95
X ² / gl	4,389	< 3,00
GFI	0,945	> 0,95
AGFI	0,925	> 0,95
RMSEA	0,060	< 0,08

Así mismo se ha calculado el coeficiente de fiabilidad de cada uno de los tres subconjuntos de ítems. El resultado obtenido es de 0,812 para los cinco primeros ítems correspondientes al factor Planificación, de 0,906 para los de Proceso y de 0,840 para los de Resultados.

A la vista del ajuste del modelo planteado, se ha analizado el peso de cada uno de estos factores en la explicación de la variable criterio mediante el análisis de los ajustes obtenidos con el modelo de regresión en cada una de las áreas de conocimiento.

El resumen del orden y de la bondad del ajuste obtenido (coeficiente de determinación en cada campo científico) es el que figura en la tabla 8. Conviene observar que, en 4 de las 5 áreas de conocimiento, el factor que mayor peso tiene en la explicación de la variable Criterio “En general pienso que es un buen profesor o profesora” son los ítems relativos al “Proceso” en el que se incluye la capacidad de Comunicación. Sólo en las Enseñanzas Técnicas se prioriza lo relativo a la “Planificación”. También cabe reseñar que el porcentaje de varianza explicado en las distintas

áreas de conocimiento es muy similar (en torno al 60%), siendo el área de las Enseñanzas técnicas el campo en el que se explica una menor proporción de variabilidad.

Estos resultados son básicamente coincidentes con los obtenidos por Pepe y Wang (2012) los cuales, usando técnicas de minería de datos para detectar las cualidades de un buen docente concluyen que los aspectos que más pesan en dichas percepciones se refieren a la buena organización del contenido y a un proceso comunicativo claro y ordenado. En la misma línea, los ya citados Otani, Kim y Cho (2012) analizan la efectividad relativa de los ítems concluyendo que los aspectos que más pesan en los estudiantes a la hora de valorar la actividad de los docentes son una explicación clara, el uso efectivo del tiempo de clase, la creación de un ambiente de aprendizaje positivo y el empleo adecuado de materiales bien organizados. En un estudio acerca del perfil del docente universitario ideal, San Martín, Santamaria, Hoyuelos, Ibáñez y Jerónimo (2014) encuentran factores muy similares como integradores de dicho perfil.

Tabla 7 - Orden en que los factores se introducen en la ecuación de regresión y coeficiente de determinación. Variable dependiente: Ítem criterio

	Proceso	Resultados	Planificación
Ens. Técnicas R ²	.590 (3)	.569 (2)	.520 (1)
CC. Salud R ²	.680		.696
CC. Sociales R ²	.618	.653	
Humanidades R ²	.584	.645	
Experimentales R ²	.589		.639

Validez de criterio

Una vez analizada la fiabilidad del instrumento, su estructura dimensional y la validez de constructo, en este apartado se procede al estudio de la validez del cuestionario desde la perspectiva del criterio. Para ello, en primer lugar, se analizan las correlaciones de las respuestas a cada uno de los ítems con las dadas al ítem criterio (“En general, pienso que es un buen profesor o profesora”). Los valores se muestran en la tabla 8 donde los ítems aparecen ordenados de forma descendente en función del coeficiente de correlación policórica con el criterio y donde se han agrupado en tres bloques.

El primero está formado por los 5 ítems que más alta correlación (superior a 0,695) muestran con el criterio. El segundo está conformado por los 6 ítems que muestran una correlación media en términos relativos (entre 0,675 y 0,600). Por último, el tercer grupo está compuesto por 7 ítems con bajas correlaciones relativas con el criterio (inferiores a 0,600).

Tabla 8 - Correlación policórica de los ítems con el criterio

ítem6	0,777
ítem13	0,740
ítem11	0,718
ítem17	0,713
ítem16	0,698
ítem12	0,675
ítem18	0,673
ítem2	0,664
ítem7	0,623
ítem5	0,607
ítem14	0,603
ítem15	0,594
ítem3	0,590
ítem4	0,589
ítem8	0,508
ítem1	0,466
ítem10	0,444
ítem9	0,421

Estos datos confirman la importancia de los ítems 6, 11 y 13 revelando que para los estudiantes los pilares básicos sobre los que debe basarse una buena práctica docente son un claro y organizado proceso de enseñanza, una buena motivación y una actitud entusiasta. En el otro extremo, la correlación más baja la presenta el ítem 1 (“Proporciona una Guía del Estudiante (*programación*) útil para el seguimiento de la asignatura”).

La segunda etapa del estudio de la validez de criterio se centra en analizar

el orden y peso de cada ítem en la explicación del ítem criterio, y para ello se aplicó el método de regresión por pasos segmentando la muestra para cada

uno de los cinco grandes campos científicos. En la tabla 9 se muestra el orden de entrada de los primeros ítems en cada modelo.

Tabla 9 - Orden de entrada de los ítems en cada campo científico

	Humanidades	Ciencias experimentales	Técnicas	Ciencias de la salud	Ciencias sociales y jurídicas
1º	ítem13	ítem13	ítem6	ítem11	ítem6
2º	ítem17	ítem12	ítem13	ítem6	ítem18
3º	ítem12	ítem6	ítem17	ítem8	ítem13
4º	ítem18	ítem2	ítem16	ítem13	ítem17
5º	ítem1		ítem12	ítem3	ítem12
6º					ítem16
7º					ítem8

El ítem más relevante es el número 13 (“Transmite entusiasmo impartiendo esta asignatura”) que se incluye en los 5 campos científicos. A continuación, los ítems 6 (“Explica de manera clara y ordenada”) y 12 (“Atiende las consultas que le planteamos”) aparecen en 4 de ellos.

El ítem 17 (“He aprendido mucho cursando esta asignatura”) aparece en 3 campos. Por último, el ítem 11 (“Motiva al alumnado para que se interese por su proceso de aprendizaje”) aparece sólo en el modelo de Ciencias de la Salud, pero ocupando el primer lugar, lo que muestra la importancia que los estudiantes de estas titulaciones otorgan a esta cuestión.

Por tanto, una primera conclusión en este sentido apunta a que las cuestiones que los estudiantes consideran a la hora de valorar la buena práctica docente son básicamente las mismas en todos los grados, independientemente de la adscripción a uno de los cinco grandes campos disciplinares.

Para finalizar, se presentan los resultados del estudio diferencial realizado considerando las siguientes variables: género de los estudiantes,

nivel de dificultad de las asignaturas e interés inicial y final por las mismas.

Estudio diferencial por Género.

La primera variable diferenciadora estudiada es el género de los estudiantes, como puede observarse en la tabla 10, en todos los ítems se dan diferencias entre las medias en función del género de los estudiantes. Pero no es menos cierto que se trata de diferencias muy pequeñas y en casi ningún caso resultan significativas. Las excepciones son los ítems 4 y 9 en los que las estudiantes puntúan más alto a sus docentes en la capacidad de comunicar claramente al comienzo del curso (ítem 4) y en lo relativo a la estimulación de la colaboración y el trabajo en equipo (ítem 9).

En cualquier caso, parece que la evidencia apunta a que los patrones de respuesta de los estudiantes de uno y otro género son básicamente similares.

Tabla 10 - Diferencias de las respuestas en función del género

	Mujer			Hombre			t	Sig.
	Media	N	Desv. típ.	Media	N	Desv. típ.		
ítem1	3,83	554	,995	3,71	320	,902	1,77	,071
ítem2	3,94	571	,969	3,87	330	,886	1,11	,268
ítem3	3,60	559	1,006	3,57	327	,924	0,35	,726
ítem4	4,00	560	,915	3,85	322	,933	2,30	,021
ítem5	3,51	546	1,030	3,51	315	1,001	-0,02	,984
ítem6	3,79	566	1,094	3,78	329	1,051	0,11	,908
ítem7	3,67	553	,960	3,62	318	,932	0,78	,430
ítem8	3,69	557	,937	3,69	325	,946	0,10	,917
ítem9	3,78	540	,987	3,63	321	,957	2,16	,031
ítem10	3,72	554	1,068	3,69	323	,980	0,36	,718
ítem11	3,56	559	,998	3,52	325	,970	0,60	,545
ítem12	4,05	556	,915	4,05	320	,853	0,05	,956
ítem13	3,69	554	1,135	3,73	326	1,121	-0,42	,668
ítem14	3,58	529	1,033	3,70	305	,936	-1,69	,091
ítem15	3,52	550	1,002	3,56	312	,943	-0,58	,552
ítem16	3,70	550	,951	3,74	322	,879	-0,54	,581
ítem17	3,62	564	,998	3,64	326	,965	-0,39	,692
ítem18	3,82	551	,928	3,81	325	,855	0,19	,845
Criterio	4,00	563	1,014	3,90	327	,993	1,49	,136

Estudio diferencial por Nivel de dificultad de la asignatura

Una cuestión importante recogida en el cuestionario es la valoración que el propio estudiante hace sobre la dificultad de la asignatura. Es una cuestión que está relacionada con las percepciones y valoraciones de los

estudiantes sobre la evaluación que hacen del profesorado (Addison, Best y Warrington, 2006). En la figura 4 se muestran los perfiles de valoración de cada ítem (incluido el criterio) en función de la dificultad otorgada a la materia.

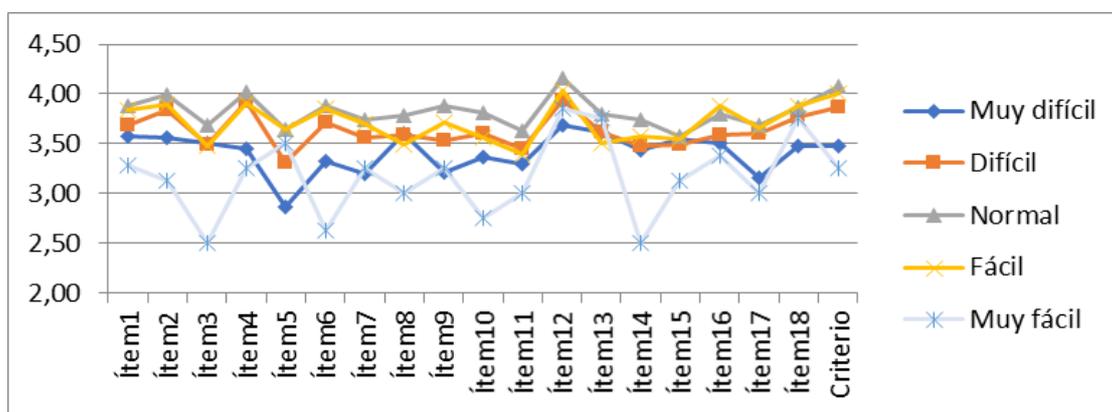


Figura 4 - Medias de los ítems según el grado de dificultad de la asignatura

En este caso, se confirma la relación apuntada en el sentido de que es perceptible una asociación entre el grado percibido de dificultad de las

asignaturas y las valoraciones que los estudiantes emiten con respecto a los docentes de las mismas. En líneas generales, la valoración más baja se

corresponde a las asignaturas consideradas como muy fáciles por los estudiantes. A continuación aparecen las “muy difíciles”. La tendencia de la

valoración más positiva corresponde a los docentes de aquellas asignaturas que son percibidas de dificultad intermedia (tabla 11).

Tabla 11 - Medias y desviaciones típicas de los ítems según nivel de dificultad de la asignatura

	Media	N	Desv. típ.
Muy difícil	3,48	44	1,267
Difícil	3,87	269	1,011
Normal	4,07	517	,954
Fácil	4,00	70	1,063
Muy fácil	3,25	8	1,282

Estudio diferencial por Interés inicial y final por la asignatura.

Otra cuestión importante es el interés que el estudiante declara por la asignatura cuyo docente está evaluando. Y al respecto, el estudiante ha de valorar el interés al inicio del curso y al final.

Para estudiar esta cuestión, lo primero que se realizó fue una comparación de la valoración de cada ítem en función del interés inicial por la asignatura. Como se observa en la figura 5, en todos los ítems se detecta la misma tendencia: a mayor interés inicial en la asignatura, mayor valoración en todos los aspectos recogidos en el cuestionario.

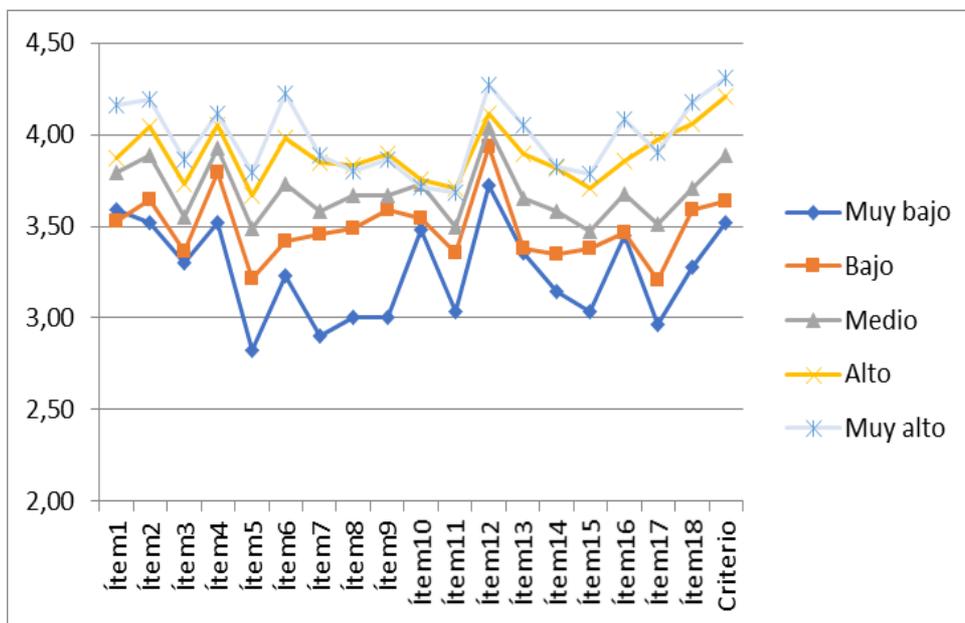


Figura 5 - Medias de los ítems según el interés inicial por la asignatura

Considerando que, como cabía esperar, las valoraciones del interés inicial y final por las asignaturas son

variables correlacionadas (0,535), se procedió a generar una nueva variable denominada *Incremento del interés*

hacia la asignatura. Para su cálculo se ha obtenido la diferencia entre el interés final y el inicial. Esta variable puede tomar valores negativos (disminuye el interés por la asignatura) y positivos (aumenta el interés por la asignatura). De esta manera se han comparado las valoraciones de cada ítem en las personas con incrementos positivos, con aquellas que tienen incrementos negativos y las que no han variado en su interés.

Los valores obtenidos se reflejan en la tabla 12. Se puede apreciar que en más de la mitad de los casos (58,5%) no ha variado el interés por la asignatura, en el 16,3% ha disminuido o empeorado y en el 25,3% ha aumentado o mejorado.

Tabla 12. Cambios en el interés por la asignatura

	Frecuencia	Porcentaje válido
Disminuye interés	150	16,3
Igual interés	539	58,5
Aumenta interés	233	25,3

En la figura 6, se pueden observar los promedios de cada ítem en función de las tres categorías (disminución, sin cambios y aumento) establecidas en función de la diferencia entre el interés inicial y final. En el mismo es perceptible la relación existente entre el incremento del interés y la valoración del docente. La puntuación media de los ítems para los casos en que el interés ha disminuido es más baja que cuando el interés se mantiene constante. Y, también para todos los ítems, las medias más altas se dan cuando el o la docente ha conseguido incrementar el interés por la asignatura.

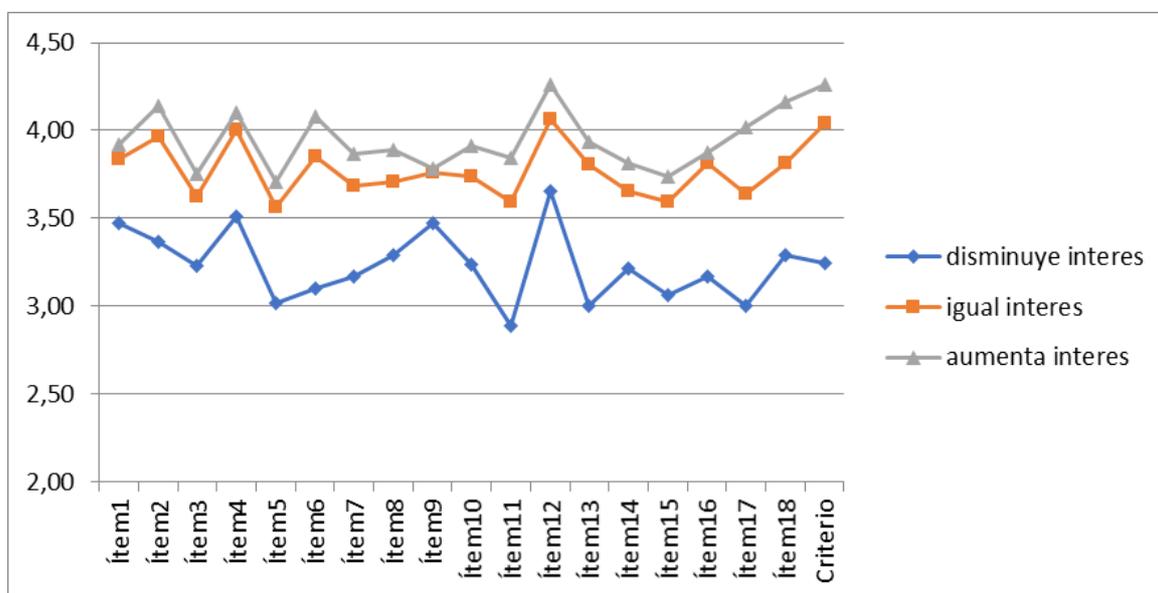


Figura 6: Medias de los ítems según los cambios en el interés por la asignatura

Se puede asegurar que una buena práctica docente consigue aumentar en los estudiantes el interés por la asignatura, lo que se traduce en un incremento de las valoraciones.

Discusión

A continuación se presentan las principales conclusiones de este estudio. En lo que se refiere a las propiedades métricas del instrumento, hay que

señalar que se trata de un cuestionario de alta consistencia interna. Los valores del Alfa de Cronbach son elevados y no hay ningún ítem cuya eventual eliminación del conjunto de la prueba pudiese implicar un incremento relevante del valor. Asimismo, las correlaciones entre los elementos y el total son elevadas.

Los resultados son congruentes por los apuntados por Apodaca y Grad (2002, 2005) relativos a una versión anterior del cuestionario de la UPV/EHU y también a los estudios ya reseñados previamente de Mortelmans y Spooren (2009) y de Burdsal y Harrison (2008). En todos los casos los resultados apuntan en una misma dirección: en este tipo de instrumentos subyace una dimensión de eficacia, profesionalidad o “buenas prácticas” docentes.

Sin embargo, lo anterior no es óbice para que al analizar esos patrones de respuesta agrupándolos en las tres dimensiones teóricamente establecidas, los resultados empíricos no difieran significativamente de dicha estructura teórica. Esto permite realizar una evaluación formativa de la calidad docente distinguiendo entre los diferentes aspectos considerados, a saber, la planificación, el proceso y los resultados.

Y al respecto no ha de olvidarse que lo importante es el uso que de esta información se haga. El debate entre los enfoques sumativo y formativo de este tipo de evaluaciones cobra aquí todo su sentido, así como las recomendaciones de Berk (2013) relativas a la posibilidad de emplear un único ítem global para tomar decisiones sumativas (como al ítem criterio al que aquí se ha hecho referencia), frente a la capacidad formativa de las diferentes agrupaciones de ítems aportando feedback al docente.

Al margen de lo anteriormente afirmado sobre la estructura dimensional del instrumento, del estudio de validez se puede concluir que una explicación de la materia de una manera clara y ordenada, el logro de un adecuado nivel de motivación e interés y una atención correcta a las consultas formuladas por los estudiantes son los tres ejes principales sobre los que parece estructurarse, lo que -a juicio de los propios estudiantes- conforma una buena praxis docente. La cuestión de las competencias profesionales no es aún conocida o valorada suficientemente. Y estas cuestiones son básicamente las mismas en todos los grados, independientemente de la adscripción de cada uno a uno de los cinco grandes campos disciplinares.

Con respecto al estudio diferencial, en lo referente a las posibles diferencias en función del género, a nivel general no hay diferencias entre las valoraciones efectuadas por las estudiantes y los estudiantes.

La dificultad de la asignatura es otra variable estudiada y los resultados apuntan a que hay una asociación entre el grado percibido de dificultad de las materias y las valoraciones que los estudiantes emiten con respecto a los docentes de las mismas. En líneas generales, la valoración tiende a ser más baja cuanto más difícil es considerada la asignatura excepto en el caso de las asignaturas consideradas como muy fáciles para las que las valoraciones disminuyen.

Es de reseñar que a mayor interés inicial en la asignatura, mejor valoración en todos los aspectos recogidos en el cuestionario. Hay una clara relación entre el incremento del interés y la valoración del docente. La puntuación media de los ítems para los casos en que el interés ha disminuido es en todos los casos más baja que cuando

el interés se mantiene constante. Y, también para todos los ítems, las medias más altas se dan cuando el o la docente ha conseguido incrementar el interés por la asignatura.

En este caso es probable que nos encontremos ante una buena práctica docente que consigue aumentar en los estudiantes el interés por la asignatura, lo que se traduce en un incremento de las valoraciones del estudiante hacia el docente.

Como conclusión final, los resultados obtenidos apuntan a que este cuestionario puede ser una herramienta útil para los fines para los que ha sido diseñado. De la misma manera, posee unas propiedades psicométricas adecuadas. Por otro lado, permite proporcionar un adecuado feedback al docente por lo que es una herramienta valiosa para la mejora de la calidad docente. Así mismo, este cuestionario muestra un grado elevado de adaptación a los nuevos modelos y estrategias de enseñanza-aprendizaje derivados del Espacio Europeo de Educación Superior (EEES).

Referencias

- Adams, Meredith J.D., & Umbach, Paul D. (2012). Non response and online student evaluations of teaching: Understanding the influence of salience, fatigue, and academic environments. *Research in Higher Education*, 53(5), 576-91. doi: <https://doi.org/10.1007/s11162-011-9240-5>
- Addison, W. E., Best, J. & Warrington, J. D. (2006). Students' Perceptions of Course Difficulty and Their Ratings of the Instructor. *College student journal*, 40(2), 409-16.
- Alvarado, Elías, Morales Dionicio & Aguayo, Ernesto (2016). Percepción de la calidad educativa: caso aplicado a estudiantes de la Universidad Autónoma de Nuevo León y del Instituto Tecnológico de Estudios Superiores de Monterrey. *Revista de la Educación Superior*, 45(180), 55-74. doi: <https://doi.org/10.1016/j.resu.2016.06.006>
- Apodaca, Pedro & Grad, Héctor (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, Pedro & Grad, Héctor (2005). The dimensionality of student ratings of teaching: integration of uni-and multidimensional models. *Studies in Higher Education*, 30(6), 723-48. doi: <https://doi.org/10.1080/03075070500340101>
- Basto, Mário & Pereira, José Manuel. (2012). An SPSS R-Menu for ordinal factor analysis. *Journal of Statistical Software*, 46(4), 1-29. doi: <https://doi.org/10.18637/jss.v046.i04>
- Berk, Ronald A. (2013). Should Global Ítems on Student Rating Scales Be Used for Summative Decisions?. *The Journal of Faculty Development*, 27(1), 63-68.
- Burdsal, Charles A. & Harrison, Paul D. (2008). Further evidence supporting the validity of both a multidimensional profile and an overall evaluation of teaching effectiveness. *Assessment & Evaluation in Higher Education*, 33(5), 567-76. doi: <https://doi.org/10.1080/02602930701699049>
- Caldera, Juan F., Carranza, María del R., Jiménez, Alma A. & Pérez, Ignacio (2015). Actitudes de los estudiantes universitarios ante la tutoría. Diseño de una escala de medición. *Revista de la Educación Superior*, 1(173), 103-124. doi:

- <https://doi.org/10.1016/j.resu.2015.04.004>
- Casero, Antonio (2010). Factores moduladores de la percepción de la calidad docente. *RELIEVE*, 16(2). doi: <http://doi.org/10.7203/relieve.16.2.4135>
- Cattell, Raymond B. (1966). The Scree Test For The Number Of Factors. *Multivariate Behavioral Research*, 1, 245-276. doi: https://doi.org/10.1207/s15327906mbr0102_10
- Chen, Guo-Hai & Watkins, David. (2010). Stability and correlates of student evaluations of teaching at a Chinese university. *Assessment & Evaluation in Higher Education*, 35(6), 675-85. doi: <https://doi.org/10.1080/02602930902977715>
- Choi, Bo-Keum & Kim, Jae-Woong. (2014). The Influence of Student and Course Characteristics on Monotonic Response Patterns in Student Evaluation of Teaching in South Korea. *Asia Pacific Education Review*, may, 1-10. doi: <https://doi.org/10.1007/s12564-014-9332-y>
- Darby, Jenny A. (2007). Are course evaluations subject to a halo effect? *Research in Education*, 77(1), 46-55. doi: <https://doi.org/10.7227/RIE.77.4>
- De Juanas Oliva, Angel & Beltrán Llera, Jesús A. (2014). Valoraciones de los estudiantes de ciencias de la educación sobre la calidad de la docencia universitaria. *Educación XXI*, 17(1), 59-82. doi: <https://doi.org/10.5944/educxx1.17.1.10705>
- Fernández Rico, J. Esteban, Fernández Fernández, Samuel, Álvarez Suárez, Alberto & Martínez Cambor, Pablo (2007). Éxito académico y satisfacción de estudiantes con la enseñanza universitaria. *RELIEVE*, 13(2). doi: <http://doi.org/10.7203/relieve.13.2.4207>
- Ginns, Paul, Prosser, Michael & Barrie, Simon. (2007). Students' perceptions of teaching quality in higher education: The perspective of currently enrolled students. *Studies in Higher Education*, 32(5), 603-615. doi: <https://doi.org/10.1080/03075070701573773>
- Glorfeld, Louis W. (1995). An Improvement on Horn's Parallel Analysis Methodology for Selecting the Correct Number of Factors to Retain. *Educational & Psychological Measurement*, 55, 377-393. doi: <https://doi.org/10.1177/0013164495055003002>
- González López, Ignacio (2003). Determinación de los elementos que condicionan la calidad de la universidad: aplicación práctica de un análisis factorial. *RELIEVE*, 9(1). doi: <http://doi.org/10.7203/relieve.9.1.4351>
- Haarala-Muhonen, Anne, Ruohoniemi, Mirja, Katajavuori, Nina & Lindblom-Ylänne, Sari. (2011). Comparison of students' perceptions of their teaching-learning environments in three professional academic disciplines: A valuable tool for quality enhancement. *Learning Environments Research*, 14(2), 155-69. doi: <https://doi.org/10.1007/s10984-011-9087-x>
- Hooper, Daire, Coughlan, Joseph & Mullen, Michael (2008). Structural Equation Modelling: Guidelines for Determining Model Fit. *The Electronic Journal of Business Research Methods*, 6(1), 53-60.
- Hoyuelos, F. J.; Ibáñez, J; Jerónimo, E.; San Martín, S. & Santamaría, M. (2014). Variables definitorias del perfil del profesor/a universitario/a ideal desde la perspectiva de los estudiantes pre-universitarios/as.

- Educación XXI*, 17(2), 193-215. doi: <https://doi.org/10.5944/educxx1.17.2.1486>
- Kaplan, David (2009). *Structural Equation Modeling (2nd ed.): Foundations and Extensions*. Thousand Oaks, USA: Sage.
- Kember, David & Leung, Doris YP. (2011). Disciplinary differences in student ratings of teaching quality. *Research in Higher Education*, 52(3), 278-99. doi: <https://doi.org/10.1007/s11162-010-9194-z>
- Kline, Rex (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York, London: The Guilford Press.
- Lance, C., Butts, M. & Michels, L.. (2006). The Sources of Four Commonly Reported Cutoff Criteria What Did They Really Say? *Organizational Research Methods*, 9(2), 202-20. doi: <https://doi.org/10.1177/1094428105284919>
- Ledesma, Rubén Daniel & Valero-Mora, Pedro (2007). Determining the Number of Factors to Retain in EFA: An Easy-to-Use Computer Program for Carrying Out Parallel Analysis. *Practical Assessment, Research & Evaluation*, 12(2), 1-11.
- Lemos, M. S., Queirós, C., Teixeira, P.M. & Menezes, I. (2011). Development and validation of a theoretically based, multidimensional questionnaire of student evaluation of university teaching. *Assessment & evaluation in higher education*, 36(7), 843-64. doi: <https://doi.org/10.1080/02602938.2010.493969>
- Lukas, José Francisco, Santiago, Karlos, Etxeberria, Juan & Lizasoain, Luis (2014). Adaptación al Espacio Europeo de Educación Superior de un cuestionario de opinión del alumnado sobre la docencia de su profesorado. *RELIEVE*, 20(1), art. 2. doi: <http://doi.org/10.7203/relieve.20.1.3812>
- Molero López Barajas, David (2007). Rendimiento académico y opinión sobre la docencia del alumnado participante en experiencias piloto de implantación del Espacio Europeo de Educación Superior. *RELIEVE*, 13(2), art. 2. doi: <http://doi.org/10.7203/relieve.13.2.4205>
- Mortelmans, D. & Spooren, P. (2009). A revalidation of the SET37 questionnaire for student evaluations of teaching. *Educational Studies*, 35(5), 547-552. doi: <https://doi.org/10.1080/03055690902880299>
- Muñoz Cantero, J.M., Ríos de Deus, M.P & Abalde, E. (2002). Evaluación Docente vs. Evaluación de la Calidad. *RELIEVE*, 8(2), art. 4. doi: <http://doi.org/10.7203/relieve.8.2.4362>
- Otani, Koichiro, B., Joon Kim & Jeong-IL Cho (2012). Student evaluation of teaching (SET) in higher education: How to use SET more effectively and efficiently in public affairs education. *Journal of Public Affairs Education*, 18(3), 531-544.
- Palmer, Stuart (2012). The performance of a student evaluation of teaching system. *Assessment & Evaluation in Higher Education*, 37(8), 975-985. doi: <https://doi.org/10.1080/02602938.2011.592935>
- Pascual Gómez, Isabel (2007). Análisis de la Satisfacción del Alumno con la Docencia Recibida: Un Estudio con Modelos Jerárquicos Lineales. *RELIEVE*, 13(1), art. 6. doi: <http://doi.org/10.7203/relieve.13.1.4216>
- Pepe, Julie W., & Wang, Morgan C. (2012). What Instructor Qualities Do

- Students Reward. *College Student Journal*, 46(3), 603-14.
- Peres-Neto, Pedro R., Jackson, Donald A. & Somers, Keith M. (2005). How Many Principal Components? Stopping Rules for Determining the Number of Non-Trivial Axes Revisited. *Computational Statistics & Data Analysis*, 49, 974-997. doi: <https://doi.org/10.1016/j.csda.2004.06.015>
- Rantanen, Pekka (2013). The number of feedbacks needed for reliable evaluation. A multilevel analysis of the reliability, stability and generalisability of students' evaluation of teaching. *Assessment & Evaluation in Higher Education*, 38(2), 224-239. doi: <https://doi.org/10.1080/02602938.2011.625471>
- Revelle, William & Rocklin, Thomas (1979). Very Simple Structure. Alternative Procedure for Estimating the Optimal Number of Interpretable Factors. *Multivariate Behavioral Research*, 14(4), 403-414. doi: https://doi.org/10.1207/s15327906mbr1404_2
- Schreider, J., Stage, F., King, J., Nora, A. & Barlow, E. (2006). Reporting structural equation modeling and confirmatory factor analysis results: a review. *The Journal of Education Research*, 99(6), 323-337. doi: <https://doi.org/10.3200/JOER.99.6.323-338>
- Spooren, Pieter, Brockx, Bert & Mortelmans, Dimitri. (2013). On the Validity of Student Evaluation of Teaching The State of the Art. *Review of Educational Research*, 83(4), 598-642. doi: <https://doi.org/10.3102/0034654313496870>
- Stark-Wroblewski, Kimberly, Ahlering, Robert F. & Brill, Flannery M. (2007). Toward a more comprehensive approach to evaluating teaching effectiveness: Supplementing student evaluations of teaching with pre-post learning measures. *Assessment & Evaluation in Higher Education*, 32(4), 403-15. doi: <https://doi.org/10.1080/02602930600898536>
- Stout, William F. (1990). A new ítem response theory modeling approach with applications to unidimensionality assessment and ability estimation. *Psychometrika*, 55(2), 293-325. doi: <https://doi.org/10.1007/BF02295289>
- Stout, William F. (1987). A nonparametric approach for assessing latent trait unidimensionality. *Psychometrika*, 52(4), 589-617. doi: <https://doi.org/10.1007/BF02294821>
- Tomkiewicz, Joseph & Bass, Kenneth. (2008). Differences between Male Students' and Female Students' Perception of Professors. *College Student Journal*, 42(2), 422-430.
- Uttl, B., White, C. A. & Gonzalez, D. W. (2016). 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. doi: <http://dx.doi.org/10.1016/j.stueduc.2016.08.007>
- Velicer, Wayne F., Eaton, Cheryl A. & Fava, Joseph L. (2000). Construct explication through factor or component analysis: A review and evaluation of alternative procedures for determining the number of factors or components. En *Problems and solutions in human assessment*, 41-71. Springer. doi: https://doi.org/10.1007/978-1-4615-4397-8_3
- Wood, James M., Tataryn, Douglas J. & Gorsuch, Richard L. (1996). Effects of under-and overextraction on principal axis factor analysis with varimax

rotation. *Psychological methods*, 1(4), 354. doi: <https://doi.org/10.1037/1082-989X.1.4.354>

Zerihun, Zenawi, Beishuizen, Jos & Van Os, Willem. (2012). Student learning experience as indicator of teaching quality. *Educational Assessment, Evaluation and Accountability*, 24(2), 99-111. doi: <https://doi.org/10.1007/s11092-011-9140-4>

Zhao, Jing & Gallant, Dorinda J. (2012). Student evaluation of instruction in higher education: Exploring issues of validity and reliability. *Assessment & Evaluation in Higher Education*, 37(2), 227-35.

doi: <https://doi.org/10.1080/02602938.2010.523819>

Zwick, William R. & Velicer, Wayne F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological bulletin*, 99(3), 432-442. doi: <https://doi.org/10.1037/0033-2909.99.3.432>

NOTA

Este trabajo se ha realizado con el apoyo del Vicerrectorado de Calidad e Innovación Docente de la Universidad del País Vasco (UPV/EHU).

Autores / Authors

To know more / Saber más

Lizasoain-Hernández, Luis. (luis.lizasoain@ehu.es).

Doctor en Filosofía y Ciencias de la Educación. Profesor Titular de Estadística en el Departamento de Métodos de Investigación y Diagnóstico en Educación (MIDE) de la Universidad del País Vasco en San Sebastián, imparte su docencia tanto en la UPV-EHU como en diversas universidades latinoamericanas. Su dirección postal es: Departamento MIDE. Facultad de Educación, Filosofía y Antropología de la UPV-EHU. Avda. de Tolosa 70. 20018-San Sebastián (España).


[0000-0002-0753-2806](https://orcid.org/0000-0002-0753-2806)



Etxeberria-Murgiondo, Juan (juanito@ehu.es).

Doctor en Ciencias de la Educación y Licenciado en Ciencias Exactas (Especialidad Estadística). Catedrático en el Departamento MIDE de la Universidad del País Vasco en San Sebastián. Ha sido subdirector de los Cursos de Verano que la Universidad del País Vasco organiza en San Sebastián. Su dirección postal es: Departamento MIDE. Facultad de Educación, Filosofía y Antropología de la UPV-EHU. Avda. de Tolosa 70. 20018-San Sebastián (España).


[0000-0003-4346-4297](https://orcid.org/0000-0003-4346-4297)

Lukas-Mujika, José F. (jf.lukas@ehu.es).

Doctor en Pedagogía, Licenciado en Filosofía y Ciencias de la Educación (sección Pedagogía) y Diplomado en Magisterio. Profesor Titular de Universidad en el Departamento MIDE de la Universidad del País Vasco. Especialista en Evaluación en Educación con especial dedicación a temas relacionados con la evaluación de programas y centros educativos y la construcción de instrumentos de medida. Es el autor de contacto para este artículo. Es editor de sección de RELIEVE. Su dirección postal es: Departamento MIDE. Facultad de Educación, Filosofía y Antropología de la UPV-EHU. Avda. de Tolosa 70. 20018-San Sebastián (España).


[0000-0002-9440-035X](https://orcid.org/0000-0002-9440-035X)





Revista ELectrónica de Investigación y EValuación Educativa
E-Journal of Educational Research, Assessment and Evaluation

[ISSN: 1134-4032]

© Copyright, RELIEVE. Reproduction and distribution of this articles it is authorized if the content is no modified and their origin is indicated (RELIEVE Journal, volume, number and electronic address of the document).

© Copyright, RELIEVE. Se autoriza la reproducción y distribución de este artículo siempre que no se modifique el contenido y se indique su origen (RELIEVE, volumen, número y dirección electrónica del documento).