



Estudio psicométrico y validación de un cuestionario para la evaluación del profesorado universitario de enseñanza a distancia

Psychometric study and validation of a questionnaire for the evaluation of online university lecturers

Isabel Cañadas*, Isabel de la Cuétara**

*Universidad Miguel Hernández de Elche, **Universidad Nacional de Educación a Distancia

Resumen

En este trabajo se presenta el estudio psicométrico del *Cuestionario de Evaluación del Profesorado Universitario versión a Distancia (CEPU, versión D)*, cuya finalidad es evaluar la docencia de los profesores que ejercen su trabajo en la modalidad no presencial, sobre la base de las consideraciones previas del estudiante. El cuestionario ha sido construido utilizando una combinación de metodología cualitativa y cuantitativa. La novedad más relevante de esta escala es que se ha creado a partir de la opinión de los estudiantes sobre las cualidades que debe tener un profesor *a distancia (online)* para ser considerado un buen profesor. En anteriores trabajos hemos presentado el procedimiento empleado para construir la escala; en este trabajo presentamos el estudio psicométrico del cuestionario específico para tutores a distancia. La muestra se seleccionó mediante muestreo aleatorio por conglomerados, considerando como conglomerado el Campus. Se llevó a cabo un análisis factorial exploratorio (AFE) y a continuación un análisis factorial confirmatorio (AFC). Los resultados indican que el cuestionario tiene una alta fiabilidad y validez, lo que lo convierte en una herramienta de evaluación útil y relevante.

Palabras clave: calidad universitaria; necesidades de los estudiantes; validez de cuestionarios

Abstract

In this paper we present the psychometric study of the *Questionnaire for the Evaluation of Online University Lecturers (version D)*, whose purpose is to evaluate the teaching of professors who work online mode, on the basis of the previous considerations of the student. The questionnaire has been constructed using a combination of qualitative and quantitative methodology. The particularly relevant feature of this scale is that it has been generated from the students' opinion about the characteristics that online lecturers must have in order to be considered a good teacher. In previous works we have presented the procedure that has been used to build the scale; in this we present the psychometric study of the specific questionnaire for online tutors. The sample was selected by random sampling by conglomerates, considering the Campus as a conglomerate. An exploratory factorial analysis (AFE) was carried out and subsequently a confirmatory factorial analysis (CFA). The results indicate that the constructed questionnaire has a high reliability and validity, which makes it a useful and relevant evaluation tool.

Keywords: university quality; students' needs; validity of questionnaire

Isabel Cañadas  orcid.org/0000-0002-7292-0426, Departamento de Psicología de la Salud, Universidad Miguel Hernández, Avda. de la Universidad s/n Edificio Altamira, 03202 Elche (Alicante), España

Isabel de la Cuétara  orcid.org/0000-0002-8737-2020, Universidad Nacional de Educación a Distancia, Centro Asociado de Tenerife, San Agustín, 30, 38201 La Laguna, España

Correspondencia relativa a este artículo: Isabel Cañadas - isabel@umh.es

El uso de cuestionarios estandarizados, cumplimentados de forma anónima, constituye la principal herramienta para recabar información del estudiante universitario. Tanto si se trata de una enseñanza presencial, como si es a distancia o semipresencial, el cuestionario cerrado tipo Likert ha monopolizado el mundo de la evaluación, convirtiéndose en el instrumento por excelencia utilizado para estos fines. Esto ha dado lugar a una vasta investigación, cuestionándose aún hoy en día la validez de sus resultados. Desde el trabajo de Kratz en 1889 sobre el estudio de la efectividad del profesor a partir de la opinión del estudiante (Carrascosa y Molero, 2005), no sólo la literatura está llena de resultados sobre satisfacción y calidad docente, sino también de resultados acerca de investigaciones sobre los propios instrumentos que se utilizan. Sin embargo, evaluar la enseñanza no es sinónimo de evaluar al profesor que la imparte. Spencer y Schmelkin (2002), a partir de una encuesta por correo a una muestra aleatoria de estudiantes, informan que, generalmente, estos se muestran participativos en los procesos de evaluación de sus profesores, sin que muestren preocupación por posibles repercusiones de valoraciones negativas al profesor, lo que nos permite asegurar que el juicio que emite el evaluador (el estudiante) está exento de prejuicios que anulen su objetividad.

En el trabajo antes citado de Spencer y Schmelkin (2002) también se pone de manifiesto que los estudiantes tienen poca confianza en que sus evaluaciones se tomen en cuenta, tanto por parte de la Institución, como por parte del profesor. En relación con esta afirmación, observaron que, según los estudiantes, proporcionar información basada en los resultados obtenidos de la encuesta es el efecto más interesante de la evaluación. Además, las expectativas que los estudiantes tienen concernientes a los resultados dan lugar a un impacto directo en la motivación para participar en el proceso. Si los estudiantes no ven una conexión entre sus esfuerzos para rellenar los cuestionarios y el conocimiento de los resultados, las evaluaciones se convierten en una tarea de rutina, lo que hace de la evaluación un proceso ineficaz (Spooren, Brockx y Mortelmans, 2013). Si, como vemos, ya es desfavorable no aportar un *feedback* a la colaboración del estudiante en la evaluación, más grave resulta no tener en cuenta su opinión a la hora de elaborar las cuestiones que se deben contemplar en la evaluación ya que, de no ser así, el contenido no se corresponde con las características que, a juicio del estudiante, tiene que tener el docente y esto sí tiene una repercusión directa sobre la validez de los instrumentos.

Aunque se han realizado algunos esfuerzos por recabar la opinión del estudiante en relación con los aspectos a valorar en el proceso, esto no suele ser la práctica habitual y, como apuntan Onwuegbuzie, Daniel y Collins (2009), aunque se pueda suponer una validez aparente en los cuestionarios, la investigación reciente ha revelado diferencias en las perspectivas que las distintas partes implicadas tienen de la calidad docente. Estas diferencias

amenazan tanto la validez como la eficacia de los instrumentos. En efecto, y en relación con la validez, en el diseño de estas herramientas raramente se tiene en cuenta la opinión del estudiante y esto podría conducir, como ya hemos adelantado, a una desmotivación en la participación y a respuestas no fiables, lo que lleva a resultados sesgados (Buchanan, 2011; Pan et al., 2009; Pozo-Munoz, Reboloso-Pacheco y Fernandez-Ramirez, 2000; Rogers y Smith, 2011; Su y Wood, 2012). En este sentido, ha habido esfuerzos como el de Barth (2008), que propone un cuestionario de cinco factores que tiene en cuenta la perspectiva del alumno, o la propuesta de Pan et al. (2009) quienes, además de la metodología cuantitativa, emplean como complemento una visión cualitativa. En la misma línea, MacDonald y Gobson (2011) emplearon un procedimiento abierto en el que las reflexiones de los estudiantes se recopilaron en un papelógrafo. Sin embargo, en estos trabajos, el objetivo no era directamente la búsqueda de información sobre la satisfacción del estudiante, sino determinar las causas del abandono la carrera en el curso del segundo año.

En relación con las características y destrezas del profesor, y no de la enseñanza, Okoro y Chukwudi (2011) identificaron, en una revisión de la literatura, un número variable de factores que incluyen los siguientes aspectos: profesionales, personales, sociales, de liderazgo y de interacción en la clase. Ya sea sobre la evaluación docente o sobre la evaluación del profesor, como señalan Spooren et al. (2013), no hay un acuerdo en la literatura acerca del número y naturaleza de las dimensiones que deben ser evaluadas. Y aún hay más, algunos autores abogan en favor de una única puntuación global y en la posibilidad de compilar un único factor basado en dimensiones comentadas (Apodaca y Grad, 2005). Desde los años 90 los investigadores debaten sobre este aspecto de la dimensionalidad y sobre si las puntuaciones en muchas dimensiones deberían ser recogidas en un único factor que represente al constructo global. Muchos autores han aportado evidencias que apoyan una competencia general (Apodaca y Grad, 2005; Burdsal y Harrison, 2008; Mortelmans y Spooren, 2009), lo que facilitaría, por un lado, utilizar los resultados de una o más dimensiones concretas cuando se trabaja sobre la mejoría de un aspecto concreto de la docencia y, por otro, obtener una puntuación global correspondiente a un factor único que representa la competencia general de enseñanza, especialmente útil cuando la finalidad es evaluar al profesor.

Por otra parte, tal y como afirman Ferguson y DeFelice (2010) uno de los enfoques más importantes y significativos en la mejora del aprendizaje viene de la mano del nuevo paradigma del *e-learning* que dominará la enseñanza del siglo XXI y donde uno de los aspectos más relevantes es, según Rothman, Romeo, Brennan y Mitchell (2011), el cambio en las relaciones profesora/alumno que pasan de ser unidireccional a bidireccional al compartir un proceso de aprendizaje recíproco del que se beneficiarán ambos actores.

¿Cómo afecta este cambio a la evaluación del profesorado en este nuevo paradigma? Si sobre la evaluación del profesorado universitario de la enseñanza presencial existen, como hemos visto, tantas opiniones, ¿qué ocurre con la enseñanza *online*?, ¿qué cabe decir de la calidad docente y de la satisfacción de los estudiantes en la enseñanza a distancia?, ¿cómo define un estudiante *online* al buen profesor?

En general, a tenor de la bibliografía revisada, parece que el estudio de la actitud del estudiante hacia el uso de las nuevas tecnologías y el *e-learning* no ha recibido mucha atención (Brown et al., 2010). Las investigaciones dirigidas al estudio de la satisfacción del estudiante *online* se han encaminado al análisis de indicadores centrados en aspectos relativos a la respuesta de la institución (hacia las necesidades del estudiante), la calidad de los contenidos *online*, la inmediatez de la retroalimentación, la frecuencia de la interacción entre el estudiante y el profesor, la disponibilidad de ayuda financiera, etc. Ferguson y DeFelice (2010) ponen de manifiesto que los aspectos estructurales del curso, la actitud hacia las nuevas tecnologías, la posibilidad de la auto administración del tiempo, la interacción con el instructor, la colaboración con otros compañeros, etc. son indicadores importantes de la calidad docente. Naaj, Nachouki y Ankit (2012) desarrollaron un instrumento para evaluar la satisfacción de los estudiantes a distancia, obteniendo cinco factores: tecnología, interacción con el profesor, dirección del curso, organización y supervisión. Hung y Chou (2014), en su cuestionario sobre el rol y la conducta del instructor *online*, obtuvieron también cinco factores: organizador y diseñador del curso, motivador, apoyador social, facilitador de tecnología y diseñador de la evaluación. Sin embargo, estos instrumentos adolecen, al igual que los utilizados en la enseñanza presencial, de no tener en cuenta para su construcción la opinión del estudiante.

En trabajos anteriores hemos presentado una metodología cualitativa que nos ha permitido determinar 15 indicadores que, a tenor de las respuestas de los estudiantes, definen al *profesor ideal* (San Luis y Cañadas, 2014). Concretamente, se utilizó el método de ordenación por rangos, combinado con el método conocido como lluvia de ideas, que dio lugar a 744 respuestas, tras lo cual, mediante revisión teórica y grupos de discusión, se obtuvieron las 16 cualidades deseables en un profesor universitario por orden de importancia para el alumno. La obtención de los valores escalares se realizó mediante el método de ordenación por rangos para muestras grandes. Sin embargo, la metodología cualitativa empleada, cuya ventaja más relevante es el incremento de la validez de la que, como ya hemos comentado, adolecen en general los cuestionarios tradicionales, presenta el grave inconveniente de un alto costo en recursos temporales y económicos, entre otras razones, por la dificultad que conlleva la cuantificación de respuestas de naturaleza cualitativa. Tampoco debemos perder de vista que la escala tipo Likert sigue siendo la

estrella, como método de evaluación, por todas las ventajas que comporta de tiempo, recursos y posibilidades de análisis.

En este trabajo presentamos el estudio psicométrico del *Cuestionario de Evaluación del Profesorado a Distancia CEPU (versión D)*. El cuestionario fue construido aunando las ventajas de las técnicas cualitativas y de las cuantitativas de un modo parsimonioso, que han dado como resultado una escala tipo Likert (*CEPU versión D*) construida a partir de las respuestas a cuestionarios abiertos sobre las cualidades que debe reunir un profesor universitario a distancia en su labor docente (Cañadas, 2013; Cañadas, Cuétara y San Luis, 2014; San Luis y Cañadas, 2014).

Método

Participantes

Sobre la población de estudiantes de Grado de la UNED se llevó a cabo un muestreo aleatorio por conglomerados considerando como conglomerado el Campus. Dentro de cada Campus los sujetos se seleccionaron por muestreo aleatorio sistemático.

Para la calcular el tamaño muestral se consideró la media como parámetro a estimar, ya que se conoce la varianza poblacional ($\sigma^2 = 95.86$ –dato obtenido en el portal de Estadística de la UNED: www.uned.es). El tamaño de la población era de 134 047 estudiantes de Grado en el curso académico 2016/17 (curso en el que se recogieron los datos de esta investigación), de los cuales el 44.5% eran varones y 55.5% mujeres. Para la determinación del tamaño muestral se fijó el grado de precisión en 1%, el intervalo de confianza en 99% y una proporción esperada de pérdidas del 30%. Con estos requisitos, el tamaño muestral es 904 sujetos. El total de unidades de análisis se distribuyó en cada Campus de forma proporcional a su tamaño. La Tabla 1 muestra la composición de cada una de las muestras de estudio por Campus.

El proceso de selección de las unidades de análisis fue por muestreo aleatorio sistemático a partir del censo conocido. El coeficiente K de elevación para esta investigación fue: $N/n = 147$. Se tomó al azar a un elemento del censo (de cada Campus) y se fueron seleccionando los demás según la secuencia $\{i, i + 1, i + 2, \dots, i + (n - 1)k\}$.

Respondieron 768 estudiantes (siendo la muestra necesaria 633 unidades si se eliminaran el 30 % de la proporción esperada de pérdidas). La edad media fue de 38.54 años ($DT = 10.8$), siendo el 37.3% hombres y el 62.7% mujeres. La amplitud de la muestra nos ha permitido dividirla en dos aplicando el criterio de (Snook y Gorsuch, 1989) para trabajos instrumentales, según el cual, con 300 participantes se obtienen soluciones fiables. De esta forma, podemos realizar el estudio de validación cruzada al tener datos de dos muestras de la misma población. La división de la muestra inicial en dos mitades aleatorias se realizó mediante la función *selección de casos* del SPSS.

Tabla 1
Tamaño muestral por Campus

CAMPUS	N total	%	N muestra
Canarias	6 728	5	46
Centro-Este	6 022	4.5	41
Este	17 013	12.7	115
Madrid	26 199	19.5	178
Nordeste	12 490	9.3	85
Noroeste	15 440	11.5	105
Norte	18 102	13.5	123
Sureste	13 749	10.3	93
Suroeste	18 304	13.7	124
Total	134 047		910
Mujeres		55.54	
Varones		44.46	

La submuestra A se empleó para el estudio y selección de los ítems y para la determinación de la estructura factorial. La submuestra B para la validación cruzada mediante un análisis factorial confirmatorio (Anderson y Gerbing, 1988; Brown, 2006).

Instrumento

El *Cuestionario de Evaluación del Profesorado Universitario versión a Distancia (CEPU, versión D)* es un cuestionario desarrollado sobre la base de las 1406 respuestas obtenidas en estudios anteriores. Para la composición se seleccionaron las 45 cualidades con la mayor frecuencia de emisión, dentro de su indicador. La escala generada sigue un formato con 45 ítems de cinco alternativas de respuesta, en la que 1 significa *menos importante* y 5 *más importante*. Por ejemplo, cuando se pregunta al estudiante por la cualidad *Alegre*, este debe marcar una de las cinco casillas numeradas del 1 al 5 según el grado de importancia que le otorgue. Este procedimiento se sigue con los 44 ítems restantes.

Se compuso un cuadernillo formado por el CEPU (versión D) y las casillas correspondientes a los datos demográficos (edad, el sexo, Centro, Grado cursado y curso).

Procedimiento

A los estudiantes seleccionados se les envió desde su Centro un correo electrónico en el que se les explicaba en qué consistía la investigación, se les solicitaba su colaboración desinteresada, así como su consentimiento para utilizar las respuestas y se les indicaba la dirección web donde debían responder.

Análisis de los datos

En primer lugar, se llevó a cabo el estudio psicométrico clásico a través del estudio de la fiabilidad (media, desviación típica, índice de discriminación y alfa de Cronbach). Mediante análisis factorial exploratorio se analizó la estructura de la escala así como la dimensionalidad, previa comprobación de la adecuación

maestral e identidad de la matriz de correlaciones. Para la valoración del modelo obtenido se emplearon los índices GFI y RMSEA; para el estudio de validación cruzada del modelo se realizó un análisis factorial confirmatorio y para evaluar el ajuste del modelo se emplearon el χ^2 Satorra-Bentler, el GFI, AGFI y el RMSEA (Hu y Bentler, 1995).

Los análisis se realizaron con los programas SPSS 22.0; Factor 7.2 (Lorenzo-Seva y Ferrando, 2013); EQS 6.3 y Excell 2010.

Resultados

Estudio psicométrico CEPU (versión D)

Esta fase del estudio se realizó sobre la submuestra A (n=378), definida en el apartado correspondiente, y se centró, en primer lugar, en el análisis psicométrico de la escala. De este modo, se calcularon los principales índices descriptivos: media de los ítems, desviación típica, índice de discriminación, así como el alfa de Cronbach. En la Tabla 2 se presentan los resultados.

Se eliminaron los cuatro ítems que no alcanzaron un índice de discriminación igual a .30, siendo el alfa de Cronbach de los 40 restantes igual a .928. A estos ítems se aplicó un análisis factorial exploratorio (AFE).

Previo al AFE se estudiaron los supuestos de adecuación muestral e identidad de la matriz de correlaciones, siendo los valores de los estadísticos respectivos: KMO = .90 y Test Bartlett's = 6149.5 (gl = 760; $p = .00001$).

El análisis de componentes principales extrajo un primer factor cuyo autovalor fue igual a 10.72, siendo el del segundo componente igual a 2.38 y el tercero 2.02. A la vista del gráfico de sedimentación de la Figura 1, se puede postular un modelo de tres factores; si bien, teniendo en cuenta el autovalor del primer factor (10.9) y del segundo (2.38), se puede también proponer un modelo de dos factores e, incluso, un modelo unifactorial (nótese que el primer autovalor es más del triple que el segundo).

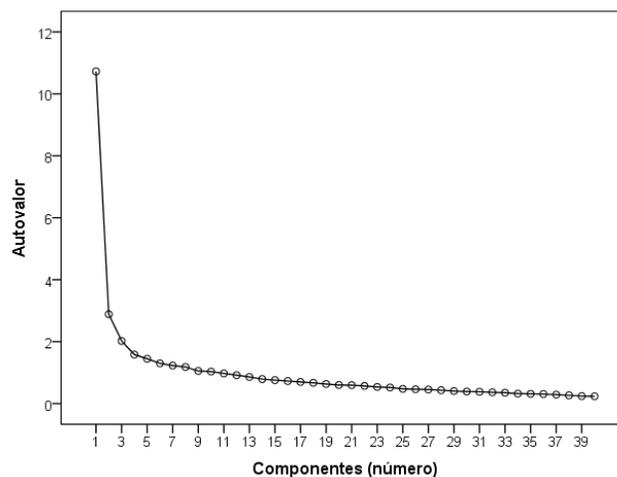


Figura 1. Gráfico de sedimentación

Tabla 2
Índices descriptivos de los 45 ítems iniciales

	Media si se elimina el ítem	Varianza si se elimina el ítem	Correlación ítem-test corregida	Alfa de Cronbach si se elimina el ítem
Alegre	172.51	334.39	.41	.92
Amable	171.23	334.39	.38	.92
Asertivo	171.46	333.27	.37	.92
Atento	171.12	333.38	.43	.92
Atractivo	174.05	340.99	.20	.93
Buen comunicador	170.69	342.31	.21	.93
Capacidad de dirigir el grupo	171.11	331.04	.51	.92
Carismático	172.16	327.28	.48	.92
Cercano	171.41	330.31	.48	.92
Claro en sus explicaciones	170.57	343.78	.23	.93
Comprensivo hacia el alumno	171.19	332.35	.44	.92
Con amplios conocimientos	170.91	337.90	.33	.93
Capacidad de síntesis	170.99	336.81	.39	.92
Con recursos didácticos	170.91	335.35	.43	.92
Creativo	171.55	327.29	.55	.92
Didáctico	170.97	335.98	.40	.92
Dinámico	171.51	325.48	.64	.92
Disponible	171.25	330.52	.50	.92
Divertido	172.91	326.59	.50	.92
Domina la materia	170.74	338.82	.36	.92
Elocuente	171.86	324.86	.56	.92
Empático	171.59	330.99	.46	.92
Estimulante	171.31	328.76	.52	.92
Exigente	171.76	332.04	.48	.92
Experimentado	171.64	324.83	.56	.92
Firme	172.28	326.50	.53	.92
Imparcial	171.37	335.44	.25	.93
Innovador	171.74	325.34	.59	.92
Inteligente	171.53	331.92	.43	.92
Justo	170.97	335.56	.42	.92
Mentalidad abierta	171.29	330.22	.49	.92
Motivado	171.05	331.98	.51	.92
Metódico	171.78	332.05	.41	.92
Motivador	170.88	334.17	.48	.92
Organizado	171.17	332.24	.50	.92
Paciente	171.62	329.11	.50	.92
Práctico	171.42	329.00	.57	.92
Preparado	170.88	335.50	.49	.92
Referente para el alumno	171.43	328.93	.51	.92
Respetuoso	170.87	336.52	.40	.92
Responsable	170.99	334.01	.45	.92
Seguro de sí mismo	171.73	326.10	.53	.92
Simpático	172.61	325.63	.52	.92
Vocacional	171.45	326.89	.47	.92

Estudio de la dimensionalidad

Para determinar la dimensionalidad del modelo, y dado que carecíamos de indicios teóricos suficientes sobre esta cuestión, dadas las peculiaridades de construcción del instrumento, se emplearon dos procedimientos: análisis en paralelo (Carroll y Green, 2003; Horn, 1965; Lattin, Timmerman y Lorenzo-Seva, 2011) y evaluación de las posibles estructuras factoriales derivadas de los criterios clásicos varianza explicada y gráfico de sedimentación. Se ejecutaron tres AFE (solicitando 1, 2 o 3 factores) con el

objeto de comparar las estructuras obtenidas, utilizando, en todos ellos, el método ULS y rotación Promax y, como matriz de entrada, la de correlaciones policóricas entre los ítems, ya que están especialmente indicadas en los casos en los que el formato de respuesta es tipo Likert (Muthén y Kaplan, 1985). El *scree-test* recomendó una solución de tres factores y el análisis paralelo confirmó la solución.

Comparación entre los modelos uni, bi y trifactorial

La Tabla 3 muestra los valores de los índices de ajuste para los tres modelos en estudio.

Tabla 3
Índices de ajuste

	GFI	RMSEA	% varianza explicada
Modelo unifactorial	.93	.08	33.2
Modelo bifactorial	.96	.06	41.3
Modelo trifactorial	.97	.05	44.67

Teniendo en cuenta que el valor recomendado para considerar un buen ajuste del modelo a los datos para el índice GFI (Tanaka y Huba, 1985) debe ser mayor de .95, es evidente que, de los tres modelos propuestos, el trifactorial es el que presenta un mejor ajuste, lo cual se confirma con el valor .05 del RMSEA, valor que coincide con el de referencia igual a .05 (Harman, 1976). Este modelo, además, es que el ofrece un mayor porcentaje de varianza explicada (44.67%). Así pues, a la vista de estos resultados, el modelo de tres factores es el que representa a los datos de forma casi óptima, seguido del de dos factores, siendo el modelo unifactorial el peor de acuerdo con los estadísticos comentados.

La Tabla 4 muestra la estructura factorial rotada correspondiente al modelo de tres factores eliminados aquellos ítems con saturaciones inferiores a .30.

En el caso de saturaciones complejas (superiores a .30 en más de un factor), se seleccionó el ítem como elemento del factor en el que presentaba la mayor carga. Los tres factores están constituidos por 16, 15 y 9 ítems respectivamente:

- Factor 1: *Profesionalidad docente*. Ítems: 9, 10, 11, 12, 13, 17, 21, 22, 23, 25, 28, 29, 31, 33, 34 y 40. (Amplios conocimientos; Capacidad de síntesis; Con recursos didácticos; Creativo; Didáctico; Exigente; Experimentado; Firme; Inteligente; Motivado; Metódico; Organizado; Preparado; Vocacional.)
- Factor 2: *Liderazgo del profesor*. Ítems: 1, 3, 5, 6, 14, 16, 18, 19, 20, 24, 27, 30, 35, 38 y 39. (Alegre; Asertivo; Capacidad de dirigir el grupo; Carismático; Dinámico; Divertido; Elocuente; Empático; Estimulante; Innovador; Mentalidad abierta; Motivador; Referente para el alumno; Seguro de sí mismo; Simpático.)
- Factor 3: *Forma de relacionarse con el alumnado*. Ítems: 2, 4, 7, 8, 15, 26, 32, 36 y 37. (Amable; Atento; Cercano; Comprensivo; Disponible; Justo; Paciente; Respetuoso Responsable.)

Análisis factorial de segundo orden

Los valores de las correlaciones entre los factores obtenidos ($r_{F_1F_2} = .56$; $r_{F_1F_3} = .58$ y $r_{F_2F_3} = .56$) indican una alta proximidad entre ellos por lo que es pertinente la obtención de un factor de segundo orden mediante rotación promin en aras a una mayor parsimoniosa (Ferrando y Anguiano-Carrasco, 2010).

Tabla 4
Matriz de cargas factoriales

Variable	F1	F2	F3
V1 Alegre		.80	
V2 Amable			.73
V3 Asertivo		.33	
V4 Atento			.54
V5 Capacidad de dirigir el grupo		.31	
V6 Carismático		.67	
V7 Cercano			.44
V8 Comprensivo hacia el alumno			.58
V9 Amplios conocimientos	.76		
V10 Capacidad de síntesis	.59		
V11 Con recursos didácticos	.63		
V12 Creativo	.45	.38	
V13 Didáctico	.59		
V14 Dinámico		.52	
V15 Disponible			.42
V16 Divertido		.82	
V17 Domina la materia	.59		
V18 Elocuente		.55	
V19 Empático		.43	
V20 Estimulante		.48	
V21 Exigente	.52		
V22 Experimentado	.69		
V23 Firme	.50		
V24 Innovador		.42	
V25 Inteligente	.43		
V26 Justo			.57
V27 Mentalidad abierta		.32	
V28 Motivado	.40		
V29 Metódico	.65		
V30 Motivador		.30	
V31 Organizado	.46		
V32 Paciente			.47
V33 Práctico	.32		
V34 Preparado	.74		
V35 Referente para el alumno		.34	
V36 Respetuoso			.77
V37 Responsable			.68
V38 Seguro de sí mismo		.55	
V39 Simpático		.82	
V40 Vocacional	.32		

La Tabla 5 muestra los resultados del AFE de primer y segundo orden. Los pesos obtenidos en el AF de primer orden se muestran en la parte izquierda de la tabla y los correspondientes al factor de segundo orden en la parte derecha. Se utiliza el subíndice SL para indicar tanto que son los factores de segundo orden como que se han obtenido mediante la solución Schmid y Leiman (1957).

Los factores $F1_{sl}$, $F2_{sl}$ y $F3_{sl}$ son independientes. El factor general G presenta pesos medio altos en todas las variables, mientras que el peso directo en los factores específicos es bastante bajo, tan sólo seis ítems superan el 0.30 en los factores $F2_{sl}$ y $F3_{sl}$ y 5 en el $F1_{sl}$.

Tabla 5
AFE de primer orden y de segundo (Método ULS, rotación Promin)

Variable	F1	F2	F3	F1 _{SL}	F2 _{SL}	F3 _{SL}	G
Alegre		.80			.49	.11	.35
Amable			.73		.10	.37	.36
Asertivo		.33		.08	.08	.13	.34
Atento			.54	-.02	.05	.33	.42
Capacidad de dirigir el grupo		.31		.15	.22	.04	.46
Carismático		.67		-.04	.45	-.03	.41
Cercano			.43	-.11	.17	.34	.45
Comprensivo hacia el alumno			.58	-.10	.05	.42	.43
Con amplios conocimientos	.76			.41	-.05	-.09	.32
Capacidad de síntesis	.59			.30	-.10	.12	.39
Con recursos didácticos	.63			.34	-.13	.15	.44
Creativo	.45			.13	.29	.02	.50
Didáctico	.59			.37	-.09	.05	.40
Dinámico		.52		.16	.24	.13	.60
Disponible			.42	.10	-.01	.33	.49
Divertido		.82		-.21	.57	.50	.42
Domina la materia	.59			.39	-.08	-.03	.35
Elocuente		.55		.11	.370	-.02	.50
Empático		.43		-.14	.23	.27	.42
Estimulante		.48		-.03	.34	.12	.47
Exigente	.52			.25	.27	-.15	.42
Experimentado	.68			.30	.22	-.08	.52
Firme	.49			.31	.28	-.18	.47
Innovador		.42		.16	.23	.08	.54
Inteligente	.43			.35	.10	-.11	.40
Justo			.57	.13	.10	.10	.37
Mentalidad abierta		.32		.05	.20	.14	.45
Motivado	.40			.10	.07	.24	.47
Metódico	.65			.29	.00	.04	.39
Motivador		.30		.02	.11	.28	.46
Organizado	.46		.30	.35	-.10	.16	.49
Paciente			.47	-.01	.10	.33	.48
Práctico	.32			.19	.04	.24	.55
Preparado	.74			.52	-.13	.01	.49
Referente para el alumno		.34		.16	.20	.06	.47
Respetuoso			.77	.12	-.07	.28	.40
Responsable			.68	.22	-.10	.22	.45
Seguro de sí mismo		.55		.17	.34	-.08	.48
Simpático		.82		-.17	.53	.06	.45
Vocacional	.32			.16	.18	.05	.44

Análisis factorial confirmatorio

Se realizó un estudio de validación cruzada para verificar la validez del modelo obtenido mediante el AFE para lo cual se empleó la submuestra B compuesta por 390 participantes, ya definida en el apartado correspondiente. De acuerdo con el modelo derivado del AFE, se propuso para la validación cruzada un modelo de tres factores, haciendo saturar cada ítem en la dimensión definida en el AFE. Se empleó el método de mínimos cuadrados ponderados y se obtuvieron los índices globales de bondad de ajuste: RMSEA = .075 (intervalo al 90% comprendido entre .071 y .078), RMSEA = .075 (intervalo al 90% comprendido entre .071 y .078), GFI = .95 y AGFI = .94, todos ellos valores indicativos de un buen ajuste al modelo.

La Figura 2 muestra la solución estandarizada correspondiente al Modelo 1. Como se puede apreciar, la correlación entre los factores indica, como en el caso del AFE, la pertinencia de un modelo factorial de segundo orden, tal que el postulado en el AFE.

Se estudió la idoneidad de la estructura factorial de segundo orden obtenida previamente en el AFE, y definida con tres factores de primer orden y un factor global. Los valores obtenidos por los índices globales de ajuste son: RMSEA = .078 (intervalo al 90% comprendido entre .073 y .080), GFI = .96 y AGFI = .96, valores que indican la validez del modelo propuesto.

Dado que, tanto a nivel exploratorio como confirmatorio, se comprobó la existencia de un factor general de segundo orden, se puede considerar la escala

unidimensional y, en consecuencia, asignar al sujeto una puntuación directa en el test calculada como la suma de los valores de las respuestas en todas las variables y, por tanto, construir un baremo en percentiles para clasificar a los sujetos.

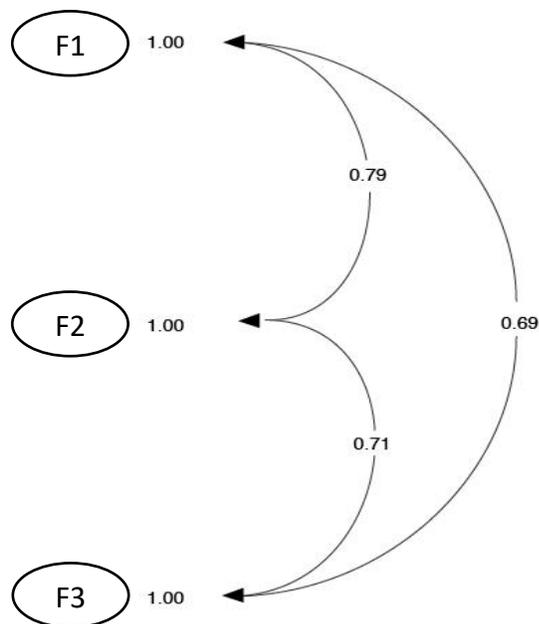


Figura 2. Modelo 1: solución estandarizada

Construcción del baremo

El procedimiento seguido para la construcción del baremo ha sido el siguiente: se excluyeron los cuatro ítems que se desecharon del análisis inicial y se calculó la

puntuación final, estudiándose la normalidad de su distribución. El histograma de esta variable se puede ver en la Figura 3, donde aparece superpuesta la curva en el caso de la normalidad. El alejamiento de las puntuaciones a esta distribución teórica fue corroborado por la prueba de Kolmogorov-Smirnov, cuyo estadístico de contraste arrojó un valor igual a .04, con $p = .001$.

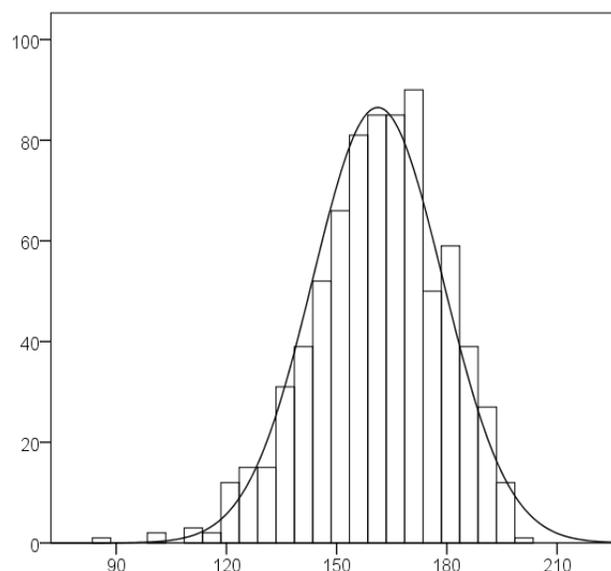


Figura 3. Distribución de las puntuaciones del cuestionario

Se realizó una doble transformación lineal de los valores extremos de la distribución de puntuaciones directas para obtener la escala de puntuaciones SE, con valores máximo y mínimo de 10 y 1, respectivamente. Los resultados de los baremos se pueden observar en la Tabla 6.

Tabla 6

Baremos en puntuaciones directas (PD), percentiles (Per) y en puntuaciones SE (SE)

PD	Per	SE	PD	Per	SE	PD	Per	SE	PD	Per	SE	PD	Per	SE
200	99	10.0	182	89	8.6	164	54	7.2	146	19	5.7	128	5	4.3
199	99	9.9	181	87	8.5	163	53	7.1	145	18	5.7	127	4	4.2
198	99	9.8	180	86	8.4	162	51	7.0	144	17	5.6	126	4	4.2
197	99	9.8	179	84	8.3	161	49	6.9	143	16	5.5	125	3	4.1
196	99	9.7	178	82	8.3	160	47	6.8	142	15	5.4	124	3	4.0
195	99	9.6	177	81	8.2	159	44	6.8	141	14	5.3	123	3	3.9
194	99	9.5	176	80	8.1	158	42	6.7	140	13	5.3	122	2	3.8
193	98	9.4	175	79	8.0	157	39	6.6	139	11	5.2	121	2	3.8
192	98	9.4	174	78	7.9	156	37	6.5	138	11	5.1	120	1	3.7
191	97	9.3	173	76	7.9	155	35	6.4	137	10	5.0	119	1	3.6
190	97	9.2	172	74	7.8	154	33	6.4	136	9	4.9	117	1	3.4
189	96	9.1	171	72	7.7	153	31	6.3	135	8	4.9	116	1	3.4
188	95	9.1	170	70	7.6	152	30	6.2	134	7	4.8	113	1	3.1
187	94	9.0	169	67	7.6	151	28	6.1	133	7	4.7	110	1	2.9
186	93	8.9	168	64	7.5	150	27	6.1	132	6	4.6	109	1	2.8
185	92	8.8	167	61	7.4	149	25	6.0	131	6	4.6	102	0	2.3
184	91	8.7	166	59	7.3	148	22	5.9	130	5	4.5	101	0	2.2
183	90	8.7	165	56	7.2	147	21	5.8	129	5	4.4	86	0	1.0

Discusión

La metodología cualitativa contribuye de forma eficaz a la recogida de respuestas de los estudiantes al concederles mucha mayor libertad de la que puedan permitir otras metodologías de carácter cuantitativo, como los cuestionarios cerrados y las encuestas, y esto tiene consecuencias directas muy favorables en la validez de contenido de los cuestionarios, como ya tuvimos ocasión de comentar en la introducción. Sin embargo, y más cuando se trata de la enseñanza superior a distancia, son los métodos cuantitativos los más parsimoniosos a la hora de recabar información de forma masiva y este hecho no puede pasar desapercibido por todas las importantes ventajas que comporta. Los resultados que obtuvimos con la metodología cualitativa nos sirvieron de base para la elaboración del CEPU (versión D) y, de este modo, hemos podido construir una escala tipo Likert, con lo que la validez aparente y de contenido del cuestionario queda garantizada.

Los 44 ítems fueron sometidos a un primer análisis exploratorio que llevó a la eliminación de cuatro ítems que no alcanzaron unos índices de discriminación iguales a .30. Estos ítems se correspondían con las cualidades *Atractivo*, *Buen comunicador*, *Claro en sus explicaciones* e *Imparcial*. El hecho de que no alcanzaran el valor mínimo de .30 bien puede deberse a que reflejan características más propias del profesor presencial, por lo que su exclusión de nuestro cuestionario no sólo obedeció a un criterio cuantitativo, sino que además se tuvo en cuenta la validez de contenido del instrumento.

Nuestros resultados del AFE indican un modelo de tres factores referidos a cualidades que tienen que ver con características tan importantes como la profesionalidad, en el sentido de que un profesor debe poseer amplios conocimientos pero, además, ser experimentado y vocacional; inteligente y organizado; didáctico y exigente y metódico y creativo (Factor 1). El docente debe ser también un líder para sus estudiantes (Factor 2): carismático, dinámico, estimulante, elocuente, motivador etc. Finalmente, debe entablar relaciones basadas en la cordialidad; características como disponibilidad, comprensión, respecto, cercanía son las que componen este factor. Estos resultados concuerdan con los encontrados en la revisión de Okoro y Chukwudi (2011) en cuanto a los tipos de atributos valorados por los estudiantes.

El análisis factorial de segundo orden dio lugar a un factor general que engloba las 30 variables que componen el cuestionario de evaluación. El modelo (trifactorial con un factor de segundo grado) quedó validado mediante AFC. El factor general obtenido ha permitido elaborar un baremo que permite clasificar a los profesores en virtud de la puntuación en su evaluación como en las propuestas de Apodaca y Grad (2005), Burdsal y Harrison (2008) y Mortelmans y Spooren (2009).

A pesar de existir muchas variedades de formas de

evaluación, la más utilizada es la encuesta dirigida al alumno; sin embargo, en general, mediante este tipo de instrumentos se estudia la labor docente desde la perspectiva institucional y desde la del profesorado, dejando de lado el enfoque propio del estudiante, es decir, sin tener en cuenta las características que él demanda de sus profesores. En efecto, en el proceso empleamos instrumentos contruidos sobre indicadores, categorías, dimensiones, etc. dadas e impuestas, sin tener en cuenta al estudiante, olvidando el componente subjetivo (estilo cognitivo, estilo de pensamiento, estilo afectivo, personalidad, nivel de formación, empatía profesor-alumno, expectativas, etc.) que afecta directamente a la evaluación. A nuestro entender la sistemática de construcción de este instrumento supera el problema antes comentado ya que son las características definidas por el estudiante las que determinan los ítems.

En este trabajo hemos identificado aspectos destacables sobre la relación profesor-alumno que no siempre forman parte de las evaluaciones mediante encuestas, utilizando un método en el que el estudiante se concibe como el elemento central de la evaluación, con plena libertad para emitir su juicio, evitando así desviar su atención hacia características más valoradas por las instituciones que por él mismo. Fruto de esa investigación ha sido el cuestionario que aquí hemos presentado, el CEPU (versión D) que posee unas propiedades psicométricas que cumplen con las exigencias metodológicas exigibles a un instrumento de medida para ser considerado válido.

Limitaciones del instrumento. Pese a la validez y fiabilidad del instrumento existen, a nuestro entender, dos limitaciones en cuanto a su uso masivo como instrumento de evaluación de la calidad del docente, ambas relacionadas con la aplicación. Por una parte, la longitud del test es una de las cuestiones que deben ser revisadas intentando acortarla en aras a evitar el cansancio del alumno durante el proceso de evaluación y mejorar las perspectivas de utilización. Por otra la forma de administración, dado que se trata de profesorado de universidades *online*, necesariamente se tiene que realizar la recogida de información a través de internet, lo que hace muy difícil obtener una muestra suficiente para considerar la evaluación válida y significativa, problema que debe resolverse motivando y haciendo al estudiante más participativo y mostrándole la relevancia de sus opiniones como sistema de mejora de la calidad de la enseñanza.

Referencias

- Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological bulletin*, 103(3), 411-423. <https://doi.org/10.1037/0033-2909.103.3.411>
- Apodaca, P., & Grad, H. (2005). The dimensionality of student ratings of teaching: integration of uni- and

- multidimensional mols. *Studies in Higher Education*, 30(6), 723-748. <https://doi.org/10.1080/03075070500340101>
- Barth, M.M. (2008). Deciphering student evaluations of teaching: A factor analysis approach. *Journal of Education for Business*, 84, 40-46. <https://doi.org/10.3200/JOEB.84.1.40-46>
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY, US: Guilford Press.
- Brown, T., Williams, B., Jaberzadeh, S., Roller, L., Palermo, C., McKenna, L. ... Sim, J. (2010). Predictors of attitudes to e-learning of Australian health care students. *Journal of Applied Research in Higher Education*, 2(1), 60-76. <https://doi.org/10.1108/17581184201000006>
- Buchanan, J. (2011). Quality teaching: means for its enhancement? *Australian Universities' Review*, 53(1), 66-72. Recuperado de <https://eric.ed.gov/?id=EJ926451>
- Burdsal, C. A., & Harrison, P. D. (2008). Further evidence supporting the validity of both a multidimensional profile and an overall evaluation of teaching effectiveness. *Assesment and Evaluation in Higher Education*, 33, 567-576. <https://doi.org/10.1080/02602930701699049>
- Cañadas, I. (2013). Study of the qualities of a good university lecturer of health sciences. *Procedia – Social and Behavioral Sciences*, 89, 342-350. <https://doi.org/10.1016/j.sbspro.2013.08.858>
- Cañadas, I., Cuétara, I., & San Luis, C. (2014). Qualities demanded by on-line universities students about their tutors. International Conference. The Future of Education. 4 Edition. 12, 13 de junio de 2014. Florence.
- Casero, A. (2010). Factores moduladores de la percepción de la calidad docente. *RELIEVE*, 16(2), 1-17. Recuperado de http://www.uv.es/RELIEVE/v16n2/RELIEVEv16n2_3.htm
- Carrascosa, J.R. y Molero, D. (2005). La valoración de la docencia universitaria en función del sexo y del centro de los estudiantes encuestados. *Revista Española de Orientación y Psicopedagogía*, 16(2), 301-318. Recuperado de <http://www.redalyc.org/pdf/3382/338230773007.pdf>
- De Juan, J, Pérez, R.M, Gómez-Torres, M.J., Vizcaya, M.F. y Mora, J.M. (2007). Buenas prácticas en la evaluación de la docencia y del profesorado universitario. En M. A. Martínez y V. Carrasco (eds.) *La multidimensionalidad de la Educación Universitaria. Redes de investigación Docente-Espacio Europeo de Enseñanza Superior* (pp. 155-182). Alcoy: Marfil. Recuperado de: <http://hdl.handle.net/10045/20334>
- Ferguson, J.M., & De Felice, A.F. (2010). Length of online course and student satisfaction, perceived learning, and academic performance. *International Review of Research in Open and Distance Learning*. 11(2), 73-84. Recuperado de <https://files.eric.ed.gov/fulltext/EJ895748.pdf>
- Ferrando, P. J. y Anguiano-Carrasco, C. (2010). El análisis factorial como técnica de investigación en psicología. *Papeles del Psicólogo*, 31(1), 18-33.
- Fu, F.L. (2010). Comparison of students' satisfaction and dissatisfaction factors in different classroom types in higher education. En Tsang P., Cheung S.K.S., Lee V.S.K. y Huang R. (eds) *Hybrid Learning*. ICHL 2010. Lecture Notes in Computer Science, vol 6248, pp. 2-38. https://doi.org/10.1007/978-3-642-14657-2_38
- Harman, H.H. (1976) *Modern factor analysis. Third edition*. Chicago: University of Chicago Press.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30(2), 179-185. <https://doi.org/10.1007/BF02289447>
- Hung, M.L., & Chou, Ch. (2014). Students' perceptions of instructors' roles in blended and online learning environments: A comparative study. *Computers & Education*, 81, 315-325. <https://doi.org/10.1016/j.comp.edu.2014.10.022>
- Hu, L. & Bentler, P. (1995). Evaluating model fit. In R. Hoyle (Ed.), *Structural equation modelling: Concepts, issues and applications* (pp.76-99). Thousand Oaks, CA: Sage.
- Kuo, Y.C., Walker, A.E., Belland, B.R., & Schroder, K.E.E., (2013). A predictive study of satisfaction in online education programs. *The International Review of Research in Open and Distance Learning*, 14(1), 16-39. <https://doi.org/10.19173/irrodl.v14i1.1338>
- Lattin, J. M., Carroll, J. D., Green, P. E., & Green, P. E. (2003). *Analyzing multivariate data*. Pacific Grove, CA: Thomson Brooks/Cole.
- Lorenzo-Seva, U., & Ferrando, P. J. (2013). FACTOR 9.2: A Comprehensive program for fitting exploratory and semiconfirmatory factor analysis and IRT models. *Applied Psychological Measurement*, 37(6), 497-498. <https://doi.org/10.1177/0146621613487794>
- MacDonald, K., & Gobson, C.E. (2011). Your tutor is your friend: using experiential learning to enhance second year transition. *Journal of Applied Research in Higher Education*, 3(2), pp.107-115. <https://doi.org/10.1108/17581181111198656>
- Martínez Arias, R. (1995). *Psicometría: Teoría de los tests psicológicos y educativos*. Madrid: Síntesis.
- Marsh, H.W. (2007). Students' evaluations of university teaching: dimensionality, reliability, validity, potential biases and usefulness. En R.P. Perry y J.C. Smart (eds.), *The scholarship of teaching and learning in higher education: An evidence-based perspective*, pp. 319-383.
- Mortelmans, D., & Spooren, P. (2009). A revalidation of the SET37-questionnaire for student evaluations of teaching. *Educational Studies*, 35, pp.547-552. <https://doi.org/10.1080/03055690902880299>
- Muthén, B., & Kaplan D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables: A note on the size of the mode. *British*

- Journal of Mathematical and Statistical Psychology*, 45, 19-30. <https://doi.org/10.1111/j.2044-8317.1985.tb00832.x>
- Naaj, M., Nachouki, M., & Ankit, A. (2012). Evaluating student satisfaction with blended learning in a gender-segregated environment. *Journal of Information Technology Education: Research*, 11(1), 185-200. Recuperado de <https://www.learntechlib.org/p/111500/>
- Okoro, C.O., & Chukwudi, E.K. (2011). The ideal teacher and the motivated student in a changing environment. *Journal of Educational and Social Research*, 1(3), 107-112. Recuperado de <http://www2.ohchr.org/english/bodies/cedaw/docs/co/NigeriaCO30.pdf>
- Onwuegbuzie, A.J., Daniel, L.G., & Collins, K.M.T. (2009). A metavalidation model for assessing the score validity of student teaching evaluations. *Quality & Quantity*, 43, 197-209. <https://doi.org/10.1007/s11135-007-9112-4>
- Pan, D., Tan, G.S.H., Ragupathi, K., Booluck, K., Roop, R., & Ip, Y.K. (2009). Profiling teacher/teaching using descriptors derived from qualitative feedback: Formative and summative applications. *Research in Higher Education*, 50, 73-100. <https://doi.org/10.1007/s11162-008-9109-4>
- Pozo-Munoz, C., Reboloso-Pacheco, E., & Fernandez-Ramirez, B. (2000). The 'Ideal Teacher'. Implications for student evaluation of teacher effectiveness. *Assessment & Evaluation in Higher Education* 25(3), 253-263. <https://doi.org/10.1080/02602930050135121>
- Rogers, J., & Smith, M. (2011). Demonstrating genuine interest in students' needs and progress: Implications for student satisfaction with courses. *Journal of Applied Research in Higher Education*, 3(1), 6-14. <https://doi.org/10.1108/17581181111150865>
- Rothman, T., Romeo, L., Brennan, M., & Mitchell, D. (2011). Criteria for assessing student satisfaction with online courses. *International Journal for e-Learning* Security, 1, 27-32. <https://doi.org/10.20533/ijels.2046.4568.2011.0004>
- San Luis, C., & Cañadas, I. (2014). Qualitative and quantitative methods to assess the qualities of a lecturer: What qualities are demanded by on-line and on-site students? *Procedia - Social and Behavioral Sciences*, 143, 106-111. <https://doi.org/10.1016/j.sbspro.2014.07.369>
- Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika*, 22, 53-61. <https://doi.org/10.1007/BF02289209>
- Snook, S. C., & Gorsuch, R. L. (1989). Component analysis versus common factor analysis: A Monte Carlo study. *Psychological Bulletin*, 106(1), 148-154. <https://doi.org/10.1037/0033-2909.106.1.148>
- Spencer, K. J. and Schmelkin, L. P. (2002). Student perspectives on teaching and its evaluation. *Assessment & Evaluation in Higher Education*, 27, 397-409. <https://doi.org/10.1080/0260293022000009285>
- 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. <https://doi.org/10.3102/0034654313496870>
- Su, F., & Wood, M. (2012). What makes a good university lecturer? Students' perceptions of teaching excellence *Journal of Applied Research in Higher Education*, 4 (2), 142-155. <https://doi.org/10.1108/17581181211273110>
- Tanaka, J.S., & Huba, G.J. (1985). A fit index for covariance structure models under arbitrary GLS estimation. *British Journal of Mathematical and Statistical Psychology*, 38, 197-201. <https://doi.org/10.1111/j.2044-8317.1985.tb00834.x>
- Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychological Methods*, 16(2), 209-220. <https://doi.org/10.1037/a0023353>

Fecha de recepción: 27 de julio de 2018.

Fecha de revisión: 19 de octubre de 2018.

Fecha de aceptación: 22 de octubre de 2018.

Fecha de publicación: 1 de diciembre de 2018.