



PROPIEDADES PSICOMÉTRICAS DE LA ESCALA DE SATISFACCIÓN DEL ALUMNADO DE LA UEX

VÍCTOR VALERO AMARO

JUAN FRANCISCO PANDURO LÓPEZ

VICERRECTORADO DE CALIDAD
Unidad Técnica de Evaluación y Calidad

BADAJOS, DICIEMBRE DE 2018

ÍNDICE

1	INTRODUCCIÓN Y RESUMEN	2
2	METODOLOGÍA DEL PRIMER ESTUDIO	3
3	ANÁLISIS DE LOS DATOS DEL PRIMER ESTUDIO.....	4
3.1	Análisis factorial exploratorio (AFE) de la escala	4
3.2	Conclusiones del AFE.....	6
3.3	Consistencia interna	8
3.4	Fiabilidad individual de los ítems	10
3.5	Fiabilidad compuesta	11
3.6	Validez convergente	11
3.7	Validez nomológica	12
4	CONCLUSIONES DEL PRIMER ESTUDIO.....	14
5	METODOLOGÍA DEL SEGUNDO ESTUDIO	15
6	ANÁLISIS DE LOS DATOS DEL SEGUNDO ESTUDIO	17
6.1	Consistencia interna	17
6.2	Fiabilidad individual de los ítems	17
6.3	Fiabilidad compuesta	18
6.4	Validez convergente	18
6.5	Validez nomológica	19
7	CONCLUSIONES FINALES	21

1 INTRODUCCIÓN Y RESUMEN

Este documento tiene como objetivo mostrar los resultados más relevantes de la evaluación de propiedades psicométricas de la nueva escala para medir la satisfacción del alumnado con la actividad docente en la UEx, que se instauró en modo piloto durante el Curso 2015/2016 y de manera general se utiliza a partir del Curso 2016/2017.

Como primer paso, se realizó una prueba previa con dicho cuestionario en una serie de asignaturas durante el segundo cuatrimestre del Curso 2015/2016. A partir de esta prueba piloto se recabaron 209 encuestas válidas de 6 asignaturas diferentes, impartidas en distintos centros de la UEx, que fueron utilizadas para un estudio exploratorio utilizando los programas IBM SPSS 22 y SmartPLS 3.0.

La principal conclusión de dicho estudio exploratorio fue que la escala es consistente, presenta una estructura adecuada y es válida para explicar la variable *satisfacción* en un porcentaje muy elevado. Con todo, se recomendaba no utilizar el indicador 9 de dicha escala como parte de la encuesta, ya que este ítem no muestra consistencia con el resto y, por tanto, no aporta información en la misma línea que los otros indicadores. Así, si la escala inicial que se había evaluado para medir la satisfacción del alumno contenía 7 indicadores (ítems 3 a 9 del cuestionario, que son aquellos que tienen impacto directo en la medida de satisfacción del alumno que se utiliza para el programa DOCENTIA), las conclusiones del estudio exploratorio recomendaban reducir dicha escala a 6 indicadores.

El indicador 9 del cuestionario que se recomendaba suprimir tenía el siguiente texto: *El docente informa sobre los resultados obtenidos de las actividades que propone*.

A partir de ahí, la UTEC de la UEx decidió no incluir el ítem 9 en la escala de satisfacción del alumno que, posteriormente, se emplea para analizar la satisfacción de los alumnos en los procesos de evaluación DOCENTIA. Así, aunque dicho ítem 9 se mantuvo en el cuestionario definitivo, no formaba parte de la escala de satisfacción del alumno y sus puntuaciones no se tenían en cuenta para el cálculo de la media de satisfacción global de un alumno con el profesorado.

Posteriormente, en el Curso 2016/2017 se analizó el cuestionario definitivo, que incluía la escala de satisfacción sin incluir el ítem 9 de manera generalizada.

Este análisis incluía los datos de la encuesta desarrollada en la UEx en el curso 2016/2017. Se contó con una muestra inicial de 38.636 cuestionarios, de los cuales finalmente fueron utilizados 26.919 casos válidos, tras el proceso de depuración y tratamiento de los datos recogidos en los distintos Centros de la UEx. La magnitud de la muestra despeja cualquier pequeña duda que pudiera plantearse respecto a la representatividad y validez de las conclusiones del estudio. Los datos fueron analizados utilizando el mismo software que en el estudio previo. Tras este análisis confirmatorio, se llegó a la conclusión que la escala actual de satisfacción que se utiliza en la UEx alcanza sobradamente las propiedades de consistencia, validez y fiabilidad. Asimismo, es un buen instrumento para medir la satisfacción del alumno pues alcanza unos valores de capacidad predictiva admisibles. Por todo ello, el estudio no plantea objeciones a que se utilice dicha escala para conocer el grado de satisfacción que los estudiantes muestran con su profesorado.

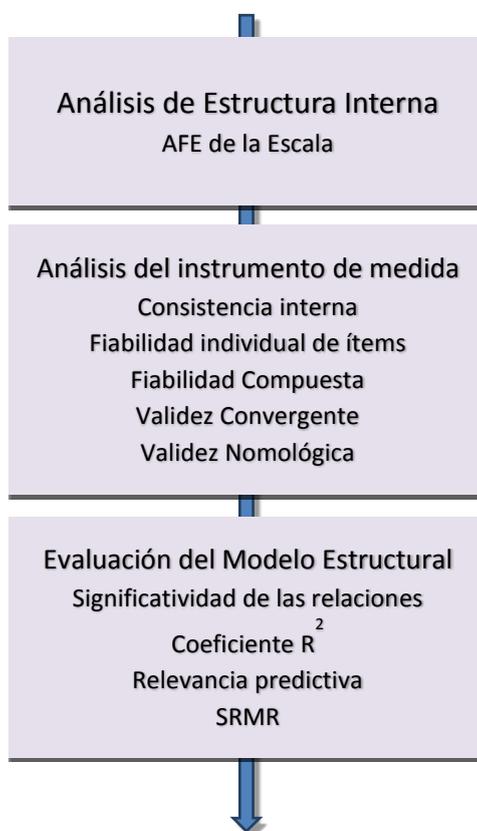
2 METODOLOGÍA DEL PRIMER ESTUDIO

La escala inicial para medir la satisfacción, de acuerdo al cuestionario elaborado por la UEx, estaba compuesta por 10 ítems o indicadores. De ellos, sólo siete componen realmente la escala de satisfacción objeto de este análisis, ya que son los indicadores 3 a 9 (ambos inclusive) los que se utilizan para determinar la puntuación de satisfacción que se utiliza, por ejemplo, en el programa DOCENTIA. De otro modo, la satisfacción del alumno que se valora para DOCENTIA se obtiene a través de una agregación de las puntuaciones que se asignan por parte de los entrevistados a esas 7 preguntas o indicadores.

En una primera etapa, se decidió realizar un análisis factorial exploratorio para la escala de satisfacción. El objetivo de este paso es analizar la estructura interna de la escala de acuerdo a los datos recogidos (209 encuestas válidas de 6 asignaturas diferentes).

Tras identificar la estructura que presenta la escala de satisfacción, se realizó el análisis del cuestionario como instrumento de medida. En este paso se evalúa su consistencia interna, fiabilidad individual de ítems, fiabilidad compuesta, validez convergente y validez discriminante. Tras el análisis de las propiedades psicométricas de la escala se evaluó su validez nomológica, lo que permite dilucidar sus propiedades para explicar y predecir la variable objeto de estudio: la satisfacción. Con ello, también se pueden añadir medidas de significatividad y ajuste. La metodología se muestra en la Ilustración 1.

Ilustración 1: Esquema de trabajo para el análisis



Fuente: Elaboración propia a partir de Sanz-Blas et al. (2008)

3 ANÁLISIS DE LOS DATOS DEL PRIMER ESTUDIO

3.1 Análisis factorial exploratorio (AFE) de la escala

Como ya se ha comentado, el análisis comienza con un estudio de su estructura interna. El objetivo es efectuar un análisis factorial exploratorio para contrastar si la estructura originalmente ideada es coincidente con la que se pueda extraer de los datos recogidos. Autores como Churchill (1979) o Gerbing y Anderson (1988) manifiestan que el análisis factorial exploratorio ha sido ampliamente utilizado en los procesos de creación de escalas, aunque su aplicación sólo puede ser satisfactoria durante las etapas tempranas de la investigación sobre un constructo (Churchill, 1979). En este sentido, se ha considerado que la creación reciente de esta escala recomienda la realización de esta investigación para ahondar en el enfoque exploratorio antes de dar pasos en el estudio confirmatorio.

Se realizó una prueba utilizando como medio de extracción el análisis de componentes principales, medio más apropiado para predecir el número de factores necesarios para justificar la porción máxima de varianza representada en la serie de variables original (Hair et al., 2007). El criterio de selección del número de factores, en este caso, ha sido el criterio de raíz latente, extrayéndose los factores con autovalores superiores a 1, es decir, explican al menos una variable. Para saturar las puntuaciones factoriales en cada factor, se aplicó la rotación VARIMAX.

La prueba se llevó a cabo utilizando las variables 3 a 9 (ambas inclusive) del cuestionario, que son aquellas que realmente tienen impacto sobre la evaluación del profesorado y que, en definitiva, componen la escala de satisfacción. No se utilizó el ítem 10 al considerarse como un indicador resumen de la información de la escala, el cual será útil para analizar la validez nomológica. No obstante, los resultados incluyendo el ítem 10 son muy similares y la inclusión de este indicador no generaba alteraciones en la estructura interna de la escala.

Una vez realizado el análisis, una observación de la matriz de correlaciones permitió comprobar que más del 60% de las correlaciones entre variables presentaba un valor superior a 0,3 y, por otro lado, prácticamente un valor similar de correlaciones resultaba estadísticamente significativa al 99,9%. Otros valores de idoneidad del análisis factorial aparecen recogidos en la Tabla 1.

Medida Kaiser-Meyer-Olkin de adecuación de muestreo		,872
Prueba de esfericidad de	Aprox. Chi-cuadrado	623,590
Bartlett	gl	21
	Sig.	,000

Tabla 1: Medidas de idoneidad del análisis

Como se puede ver en la Tabla 1, se confirma la idoneidad del análisis factorial, y tanto las medidas de Kaiser-Meyer-Olkin como la Prueba de Bartlett aportan valores satisfactorios.

De este modo, los siete indicadores forman parte del análisis factorial definitivo. El resultado de este paso extrajo 1 factor, que implicaban un 56,59% de la varianza total explicada.

Una vez alcanzado este punto, y contrastada la aplicabilidad del análisis factorial a la matriz de datos, se plantea una discusión acerca del número de factores extraídos. Como ya se ha comentado, el criterio de raíz latente extrajo un solo factor, un número consistente con la

idea de unidimensionalidad de la escala previsto a priori. Esta realidad permite plantear que todos los indicadores de la escala aportan información sobre una misma dimensión o factor, y por lo tanto tiene sentido reunirlos y agrupar su información pues es, en gran medida, convergente para conocer una única realidad. No obstante, la varianza explicada está un poco por debajo de lo habitual, pues para el campo de las ciencias sociales, es deseable que la solución final permita explicar al menos un 60% de la varianza (Hair et al., 2007).

Varianza total explicada

Componente	Autovalores iniciales			Sumas de extracción de cargas al cuadrado		
	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado
1	3,961	56,593	56,593	3,961	56,593	56,593
2	,934	13,344	69,937			
3	,698	9,971	79,908			
4	,477	6,811	86,719			
5	,344	4,912	91,630			
6	,337	4,810	96,440			
7	,249	3,560	100,000			

Método de extracción: análisis de componentes principales.

Tabla 2: Varianza explicada con autovalores mayores a 1.

La explicación de contar con una varianza un poco por debajo de lo deseable puede estar en el indicador 9 que, como se verá, es el que peor ajuste muestra y, por tanto, el único ítem sobre el que se puede plantear la discusión de su eliminación.

Este aspecto puede verse reflejado en el análisis de las comunalidades, donde se observa que el valor de la comunalidad del ítem 9 es excepcionalmente bajo. El resto de las comunalidades está en valores muy cercanos o superiores a 0'5.

Comunalidades

Extracción	
P_3	,687
P_4	,656
P_5	,717
P_6	,574
P_7	,447
P_8	,726
P_9	,156

Tabla 3: Comunalidades del AFE.

Por lo tanto, el indicador 9 ("El docente informa sobre los resultados obtenidos de las actividades que propone") parece que es el único que no aporta suficiente información sobre el factor extraído,

que, recordemos, hace referencia a la satisfacción del alumnado con la labor docente del profesorado.

Finalmente, se presentan los datos de las cargas factoriales en la matriz de componentes, donde de nuevo se pone de manifiesto que todos los indicadores de la escala presentan valores admisibles para formar parte de una misma dimensión, salvo el ítem 9.

Matriz de componente^a

	Componente 1
P_3	,829
P_4	,810
P_5	,847
P_6	,757
P_7	,668
P_8	,852
P_9	,394

Método de extracción: análisis de componentes principales.

a. 1 componentes extraídos.

Tabla 4: Matriz de componentes del AFE.

La interpretación de los factores está especialmente ligada al análisis de sus cargas factoriales, que indican la correlación que alcanza cada variable con cada uno de los factores. Inicialmente, cargas factoriales por encima del valor 0,30 indican significatividad y, por tanto, una correspondencia aceptable entre la variable y el factor. No obstante, se ha decidido adoptar un criterio más estricto para corroborar la significatividad de las cargas factoriales, aplicando el concepto de potencia estadística y buscando un nivel de potencia del 80% para un nivel de significación de 0,05. Bajo este criterio, una muestra en torno a 200 casos exige una carga factorial aproximadamente de 0,5 para garantizar su significatividad estadística. Puede observarse en la Tabla 4, por tanto, que todos los indicadores cumplen este criterio salvo el ítem 9.

3.2 Conclusiones del AFE

La principal conclusión del AFE es que los indicadores de la escala ajustan bien a un solo factor. Por tanto, sólo se extrae una única dimensión de la escala de satisfacción que permite reunir o agrupar las puntuaciones de estos ítems para alcanzar una valoración de esta variable. En otras palabras, tiene sentido hacer una valoración global de la satisfacción utilizando conjuntamente los valores de los ítems. Con todo, el ítem 9 presenta bastantes problemas y es muy probable que su eliminación de la escala mejore los valores o resultados generales de la escala.

Para confirmar este aspecto, se realizó un último AFE sin incluir el ítem 9, y se encontraron mejores resultados. Así, la escala compuesta sólo de 6 indicadores (excluyendo el ítem 9) ya permitía alcanzar más de un 60% de la varianza explicada. Los datos se muestran a continuación.

Varianza total explicada

Componente	Autovalores iniciales			Sumas de extracción de cargas al cuadrado		
	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado
1	3,734	62,242	62,242	3,734	62,242	62,242
2	,748	12,468	74,710			
3	,482	8,027	82,737			
4	,395	6,579	89,316			
5	,367	6,112	95,428			
6	,274	4,572	100,000			

Método de extracción: análisis de componentes principales.

Comunalidades

Extracción	
P_3	,701
P_4	,662
P_5	,682
P_6	,582
P_7	,406
P_8	,701

Método de extracción: análisis de componentes principales.

Matriz de componente^a

Componente 1	
P_3	,837
P_4	,814
P_5	,826
P_6	,763
P_7	,637
P_8	,837

Método de extracción: análisis de componentes principales.

a. 1 componentes extraídos.

Tablas 5, 6 y 7: Análisis AFE excluyendo el ítem 9.

De todo ello, parece desprenderse que la estructura adecuada de la escala de satisfacción de alumnado, que presenta unos valores adecuados de explicación, comunalidad y cargas factoriales, es la siguiente:

3. *El docente explica de forma clara y organizada*
4. *El docente acompaña adecuadamente las explicaciones teóricas con aspectos prácticos (ejemplos ilustrativos, casos, ejercicios, problemas,...)*
5. *El docente usa adecuadamente los recursos didácticos para favorecer el aprendizaje de la asignatura (materiales audiovisuales, campus virtual, pizarra,...)*
6. *El docente transmite motivación por el aprendizaje de la asignatura*
7. *El docente fomenta la participación activa del estudiante (expresarse en público, reflexionar, defender ideas, trabajar en grupo,...)*
8. *El docente resuelve las dudas planteadas y orienta a los estudiantes en el desarrollo de sus tareas*

La cual, además, se ajusta a un solo factor y, por tanto, nos permite percibir que presenta una estructura unidimensional, aspecto este último necesario si se desea utilizar esta escala para analizar y evaluar de manera conjunta la satisfacción con el profesorado.

3.3 Consistencia interna

Un primer elemento de análisis de la fiabilidad de una escala es el estudio de la consistencia interna de la misma. La consistencia interna, en este caso, se ha analizado a través de las correlaciones ítem- total¹ y el coeficiente alpha de Cronbach² (Cronbach, 1951), de manera individualizada por cada componente o factor (Churchill, 1979). Este estudio es complementario, para determinar una estructura interna consistente, al análisis de la varianza explicada obtenida por el factor (Churchill, 1979).

La Tabla 8 muestra los resultados del análisis de consistencia interna para la escala de satisfacción. En este proceso se han rechazado aquellos indicadores cuyas correlaciones ítem-total presentaban valores bajos (sensiblemente inferiores a 0,5), o bien su eliminación mejoraba notablemente el coeficiente Alpha. Sólo un ítem ha sido eliminado, en concreto el ítem 9, pues su correlación ítem- total no alcanzaba el valor 0,4 y, por otro lado, el coeficiente alfa de la escala mejoraba tras su rechazo.

Por otro lado, es importante destacar que el valor del coeficiente alpha de la escala definitiva (sin el indicador 9) alcanza un valor más que satisfactorio, 0'870. Por tanto, puede afirmarse que la escala de satisfacción contiene una alta consistencia interna, lo cual nos indica que todos los ítems que la forman son convergentes y muestran una información que camina en un mismo sentido. Ningún

¹ La correlación ítem-total mide la correlación que cada indicador de un constructo mantiene con la puntuación de la escala. Se recomienda que la correlación ítem-total alcance o exceda del valor 0,50 para asegurar el mantenimiento del indicador en la escala (Robinson et al., 1991).

² Los valores aceptables del coeficiente alpha de Cronbach para aceptar la consistencia interna de una escala son 0,7

(Nunnally, 1978; Nunnally y Bernstein, 1994), o incluso alcanzar el valor 0,8 (Carmines y Zeller, 1979) en el caso de escalas ampliamente utilizadas.

ítem de la escala proporciona información sobre aspectos no ligados al resto. Todo ello reafirma la percepción de unidimensionalidad de la escala.

Estadísticas de total de elemento					
	Media de escala si el elemento se ha suprimido	Varianza de la escala si el elemento se ha suprimido	Correlación total de elementos corregida	Correlación múltiple al cuadrado	Alfa de Cronbach si el elemento se ha suprimido
P_3	42,32	43,605	,730	,601	,838
P_4	41,98	45,960	,700	,554	,844
P_5	41,82	47,216	,720	,553	,844
P_6	42,41	41,758	,661	,452	,853
P_7	42,32	46,831	,526	,316	,874
P_8	41,99	43,757	,739	,568	,836

ores del análisis Alpha de Cronbach.

3.4 Fiabilidad individual de los ítems

El estudio de la fiabilidad individual, así como los restantes, se llevó a cabo utilizando la metodología de análisis de ecuaciones estructurales (SEM). En este caso, se optó por llevar a cabo análisis basados en la covarianza a través de la técnica Partial Least Squares (PLS). Se utilizó el software SmartPLS en su versión 3.0.

La fiabilidad individual se examina a través de los coeficientes *Loading* (λ) que representan las cargas o correlaciones de cada indicador con su respectivo constructo. Como regla general de fiabilidad individual, se establece que un indicador puede formar parte de una escala cuando presenta una carga igual o superior a 0,707 (Carmines y Zeller, 1979), si bien esta regla no debería ser tan rígida en las etapas iniciales del desarrollo de una escala (Chin, 1998) o cuando la escala se aplica en diferentes contextos (Barclay et al., 1995), permitiéndose indicadores con una fiabilidad inferior si su presencia mejora la validez del constructo (Hair et al., 2014) o si aportan algún a contribución a la validez de contenido.

De una primera prueba, puede observarse (Tabla 9) que el indicador 9 vuelve a mostrar un coeficiente *Loading* especialmente bajo, por lo que se recomienda de nuevo su eliminación.

	ESC_SAT
P_3	0,802
P_4	0,803
P_5	0,847
P_6	0,756
P_7	0,684
P_8	0,808
P_9	0,505

Tabla 9: Fiabilidad individual de los ítems.

Realizada otra prueba de fiabilidad individual sin el ítem 9, se alcanzan estos resultados.

	ESC_SAT
P_3	0,806
P_4	0,812
P_5	0,859
P_6	0,767
P_7	0,683
P_8	0,797

Tabla 10: Fiabilidad individual de los ítems definitiva.

Puede concluirse que, tras la eliminación del ítem 9, la escala presenta valores razonables y admisibles de fiabilidad individual.

3.5 Fiabilidad compuesta

La fiabilidad compuesta de cada escala permite comprobar, al igual que el coeficiente alpha de Cronbach anteriormente calculado, la consistencia interna de los indicadores de un constructo, es decir, su capacidad para medir conjuntamente un concepto (Barroso et al., 2005). Los valores aceptables del índice de fiabilidad compuesta (IFC o *Composite Reliability*), que indicarían una fiabilidad compuesta satisfactoria, oscilan entre 0,70 para investigaciones exploratorias y 0,80-0,90 en el caso de investigaciones más avanzadas (Nunnally y Bernstein, 1994).

El índice de Fiabilidad Compuesta (*Composite Reliability*) alcanza el valor de 0,908; pudiéndose afirmar que la escala presenta una elevada fiabilidad como instrumento de medida. Este análisis, como ya se ha comentado, es consistente con el estudio realizado anteriormente a través del coeficiente Alpha, y permite otorgar garantías a las conclusiones alcanzadas respecto a la fiabilidad de los instrumentos de medida.

3.6 Validez convergente

La validez convergente de una escala implica que el conjunto de indicadores que forman dicho constructo representan un mismo concepto subyacente (Henseler et al., 2009), es decir, todos los ítems tienden a medir una misma realidad y nada más que ella, siendo un indicador para percibir la unidimensionalidad de la escala. El análisis de la validez convergente se lleva a cabo a partir de la varianza extraída media (AVE, *Average Variance Extrated*; Hair et al., 2011), afirmándose que existe validez convergente en un constructo cuando el valor de su AVE es superior a 0,50 (Fornell y Larcker, 1981), es decir, más del 50% de la varianza del constructo es debida a sus indicadores (Barroso et al., 2005).

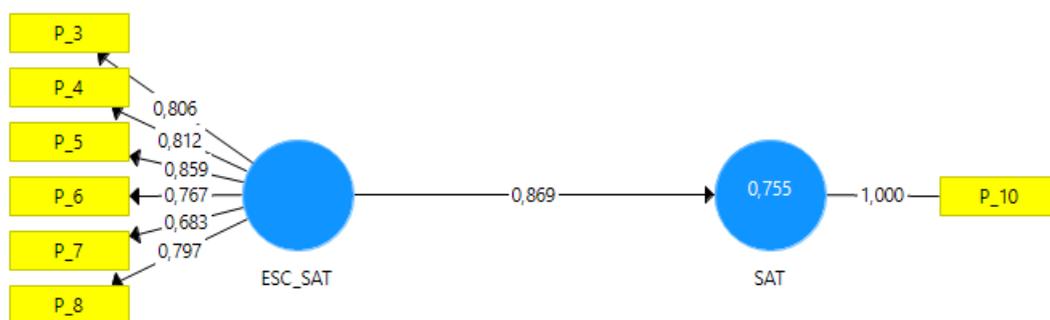
El AVE de la escala de satisfacción alcanza el valor 0,623. Se confirma que el constructo de primer orden del modelo propuesto goza de validez convergente suficiente.

3.7 Validez nomológica

Finalmente, dentro de las pruebas de validez de la escala, se analiza la validez nomológica, es decir, la capacidad de la escala para explicar o predecir un determinado concepto complejo, como es el caso de la satisfacción.

Para ello, vamos a contrastar la escala con el ítem 10 de cuestionario, que contenía información global sobre la satisfacción del alumnado. Si existe validez nomológica, la escala debería mostrar una alta correlación con el indicador 10 de satisfacción global, y además explicar un porcentaje importante de este indicador. Para ello, se ha creado un pequeño modelo estructural que relaciona la escala con este indicador global.

Ilustración 2: Modelo construido para analizar la validez nomológica



Existe validez nomológica si los indicadores de ajuste del modelo son aceptables, para ello se observa en este último paso si el ajuste de este modelo es válido.

En cuanto a la significatividad de las relaciones, el coeficiente *Path* (que analiza el camino o relación entre ambas variables) alcanza un valor de 0,869. Para contrastar si este valor es estadísticamente significativo se llevó a cabo un proceso de *Bootstrapping* con 5000 submuestras, con cuyos valores se realiza una prueba no paramétrica de significatividad del camino estructural.

El proceso de *Bootstrapping* generó un valor del estadístico *t* para el coeficiente *Path* de 42,797. Este valor es significativo al 99'99% (P-Valor = 0,000). Por tanto se puede afirmar que sí es significativo y por tanto la relación entre la escala de satisfacción y la pregunta o ítem sobre satisfacción global puede confirmarse estadísticamente.

En segundo lugar, se analiza el coeficiente R^2 . La capacidad del modelo estructural para predecir el comportamiento de las variables endógenas es evaluada por medio de los coeficientes R^2 (Chin, 1998). Este parámetro determina qué cantidad de varianza de los constructos dependientes es explicada por los constructos que la predicen (Chin, 2010). Chin (1998) especifica las fronteras de un poder predictivo bajo, medio y alto de acuerdo a que el coeficiente R^2 alcance los valores de 0'19, 0'33 y 0'67 respectivamente. Si bien en el caso de modelos estructurales que expliquen una

variable

endógena a partir de pocas variables independientes (por ejemplo, una o dos) un coeficiente R^2 bajo puede ser aceptable (Henseler et al., 2009). En este caso, el coeficiente R^2 alcanza el valor de 0,755. Es un valor muy alto, la escala de satisfacción explica un 75,5% de la varianza que presenta la pregunta de satisfacción global. Por tanto, se puede afirmar que esta escala tiene una elevada capacidad de explicación de la satisfacción del alumnado.

Otro aspecto evaluable del modelo estructural está relacionado con su relevancia predictiva, en otras palabras, la capacidad del modelo para predecir el comportamiento de las variables endógenas. La medida predominante para evaluar la relevancia predictiva es el test Q^2 de Stone-Geisser (Stone, 1974; Geisser, 1975) que analiza si el modelo es capaz de predecir adecuadamente los indicadores de los constructos endógenos (Henseler et al., 2009). El valor del estadístico Q^2 se obtiene en PLS a través de un procedimiento denominado *Blindfolding*, técnica de reutilización de la muestra que omite parte de la misma y usa el resultado estimado para predecir la parte omitida (Hair et al., 2011). El test Q^2 da un valor para nuestro caso de 0,737.

De manera general, se admite que si el valor de Q^2 es positivo, el modelo goza de capacidad para predecir el comportamiento de dicha variable (Chin, 2010; Hair et al., 2011). Del mismo modo, mayores indicadores de este parámetro sugieren mejor capacidad predictiva, pudiéndose establecer que los valores 0'02, 0'15 y 0'35 revelan una pequeña, media o elevada relevancia predictiva sobre una determinada variable latente, respectivamente (Henseler et al., 2009). Esto nos lleva a concluir que la escala mantiene una elevada capacidad para predecir la satisfacción del alumnado.

Finalmente, se analiza el coeficiente SRMR como medida de bondad del ajuste del modelo. El valor que alcanza es 0,08. Una evaluación conservadora de este índice considera que se está ante un buen ajuste cuando el índice SRMR se encuentra por debajo de 0,08 (Hu y Bentler, 1999). De acuerdo a ello, el ajuste del modelo alcanza valores deseables, aun no siendo un buen valor de ajuste.

Con todo ello, finalmente los valores del modelo causal son aceptables, y se puede concluir que la escala mantiene validez nomológica para explicar el comportamiento de la variable *Satisfacción del alumnado*.

4 CONCLUSIONES DEL PRIMER ESTUDIO

Como ya se ha ido indicando, el objetivo de este estudio preliminar era presentar unos primeros resultados acerca de las propiedades psicométricas de la nueva escala de satisfacción del alumnado. Se trata de un estudio con unos datos piloto, por lo que con posterioridad se hará un nuevo análisis con mayor profundidad con datos obtenidos de las oleadas de encuestas que se lleven a cabo.

No obstante, los datos parecen indicar que la encuesta es consistente, presenta una estructura adecuada y es válida para explicar la variable satisfacción en un porcentaje muy elevado. Con todo ello, se recomienda no utilizar el indicador 9 como parte de la encuesta, ya que este ítem no muestra consistencia con el resto y por tanto no aporta información en la misma línea que los otros indicadores.

5 METODOLOGÍA DEL SEGUNDO ESTUDIO

Como ya se ha comentado, a partir de las conclusiones del primer estudio realizado sobre la escala de satisfacción del alumnado, la UTEC de la UEx decidió no incluir el ítem 9 en dicha escala que, posteriormente se emplea para analizar la valoración de los alumnos sobre los profesores que se someten a los procesos de evaluación DOCENTIA. Por tanto, aunque dicho ítem 9 se mantuvo en el cuestionario definitivo, no formaba parte de la escala que mide la puntuación de satisfacción global del alumno.

Así, en el Curso 2016/2017 se empleó el nuevo cuestionario, que incluía la nueva escala de satisfacción (sin incluir el ítem 9) de manera generalizada.

Este segundo análisis tiene por objetivo realizar un estudio confirmatorio de las características y propiedades psicométricas del cuestionario. Para ello se han utilizado los datos de la encuesta desarrollada en la UEx en el curso 2016/2017. Se ha contado con una muestra inicial de 38.636 cuestionarios, de los cuales finalmente han sido utilizados 26.919 casos válidos, tras el proceso de depuración y tratamiento de los datos recogidos en los distintos Centros de la UEx.

Al tratarse de un análisis confirmatorio, se obvia la realización de un análisis factorial, pues la estructura interna de la escala ya fue determinada con el estudio anterior. De hecho, uno de los objetivos de este análisis es confirmar dicha estructura interna con datos reales de una oleada del cuestionario.

Por tanto, el análisis se centrará en determinar las propiedades del instrumento de medida, en términos de consistencia interna, fiabilidad individual de ítems, fiabilidad compuesta, validez convergente y validez discriminante. Tras el análisis de las propiedades psicométricas de la escala se evalúa su validez nomológica, lo que permite dilucidar sus propiedades para explicar y predecir la variable global objeto de estudio: la satisfacción. Con ello, también se pueden añadir medidas de significatividad y ajuste. La metodología se muestra en la Ilustración 3.

Ilustración 3: Esquema de trabajo para el análisis



Fuente: Elaboración propia a partir de Sanz-Blas et al. (2008)

6 ANÁLISIS DE LOS DATOS DEL SEGUNDO ESTUDIO

6.1 Consistencia interna

Los métodos de análisis de este segundo estudio son similares a los practicados en el estudio preliminar, tanto en lo concerniente a consistencia interna como a las medidas de fiabilidad y validez.

La Tabla 11 muestra los resultados del análisis de consistencia interna para la escala de satisfacción. En este proceso se rechazarían aquellos indicadores cuyas correlaciones ítem-total presenten valores bajos (sensiblemente inferiores a 0,5), o bien su eliminación mejoraba notablemente el coeficiente alpha. Ningún ítem de la escala debería ser eliminado. Las correlaciones totales no sólo son superiores a 0,5; lo son en muy alta medida (la correlación más baja es 0,791). Todos los indicadores de la escala convergen y se correlacionan con la escala, por lo que tienden a medir una misma realidad. Ningún indicador es divergente con el resto. Además, la escala no mejoraría sus valores de consistencia si se eliminara algún ítem.

Por otro lado, es importante destacar que el valor del coeficiente alpha de la escala alcanza un valor mucho más que satisfactorio: 0,940 (el valor del coeficiente alfa en el estudio exploratorio fue de 0,870). Por tanto, puede afirmarse que la escala de satisfacción contiene una alta consistencia interna, lo cual nos indica que todos los ítems que la forman son convergentes y muestran una información que camina en un mismo sentido. Ningún ítem de la escala proporciona información sobre aspectos no ligados al resto. Todo ello reafirma la percepción de unidimensionalidad de la escala.

Estadísticas de total de elemento					
	Media de escala si el elemento se ha suprimido	Varianza de escala si el elemento se ha suprimido	Correlación total de elementos corregida	Correlación múltiple al cuadrado	Alfa de Cronbach si el elemento se ha suprimido
P3	38,39	101,078	,833	,718	,927
P4	38,14	101,715	,838	,727	,926
P5	38,32	101,352	,783	,639	,933
P6	38,69	94,645	,851	,737	,925
P7	38,53	99,895	,791	,654	,932
P8	38,09	101,592	,825	,689	,928

Tabla 11: Valores del análisis alpha de Cronbach.

6.2 Fiabilidad individual de los ítems

El estudio de la fiabilidad individual, así como los restantes, se llevó a cabo utilizando la metodología de análisis de ecuaciones estructurales (SEM). Al igual que en el estudio previo, se optó por llevar a cabo análisis basados en la covarianza a través de la técnica Partial Least Squares (PLS). Se utilizó el software SmartPLS en su versión 3.0.

Puede observarse (Tabla 12) que todos los coeficientes *Loading* de la escala presentan una carga superior a 0,7. Esto nos indica que no se recomienda eliminar ningún ítem de la escala, ya

que todas las preguntas de la escala alcanzan un elevado valor de fiabilidad individual. Los ítems de la escala son fiables.

	ESC_SAT
P_3	0,890
P_4	0,837
P_5	0,781
P_6	0,900
P_7	0,807
P_8	0,887

Tabla 12: Fiabilidad individual de los ítems.

6.3 Fiabilidad compuesta

El índice de Fiabilidad Compuesta (Composite Reliability) alcanza el valor de 0,940 (el valor de este parámetro en el estudio anterior fue 0,908). Además, se ha calculado el coeficiente ρ de Spearman (se recomienda un coeficiente rho superior a 0,7) que alcanza el valor 0,942. Se puede afirmar que la escala presenta una elevada fiabilidad como instrumento de medida. Este análisis, como ya se ha comentado, es consistente con el estudio realizado anteriormente a través del coeficiente alpha, y permite otorgar garantías a las conclusiones alcanzadas respecto a la fiabilidad de los instrumentos de medida.

6.4 Validez convergente

La validez convergente de una escala implica que el conjunto de indicadores que forman dicho constructo representan un mismo concepto subyacente (Henseler et al., 2009), es decir, todos los ítems tienden a medir una misma realidad y nada más que ella, siendo un indicador para percibir la unidimensionalidad de la escala. También se denomina validez de constructo. El análisis de la validez convergente se lleva a cabo a partir de la varianza extraída media (AVE, *Average Variance Extrated*; Hair et al., 2011), afirmándose que existe validez convergente en un constructo cuando el valor de su AVE es superior a 0,50 (Fornell y Larcker, 1981), es decir, más del 50% de la varianza del constructo es debida a sus indicadores (Barroso et al., 2005).

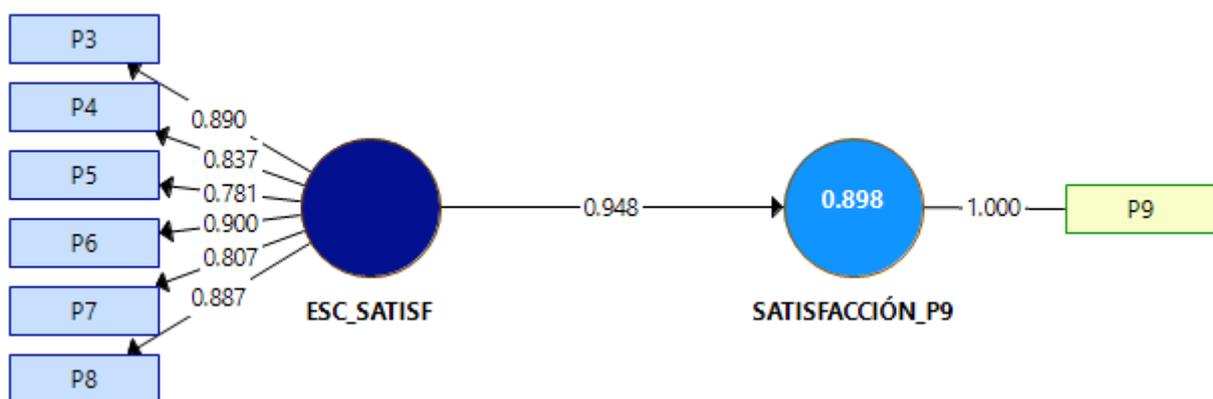
El AVE de la escala de satisfacción alcanza el valor 0,725 (0,623 fue el valor en el estudio anterior). Se confirma que el constructo de primer orden del modelo propuesto goza de validez convergente suficiente. 0,725 es un valor muy satisfactorio para el AVE de la escala.

6.5 Validez nomológica

Finalmente, dentro de las pruebas de validez de la escala, se analiza la validez nomológica, es decir, la capacidad de la escala para explicar o predecir un determinado concepto complejo, como es el caso de la satisfacción.

Para ello, vamos a contrastar la escala con la pregunta 9 del cuestionario final, que contenía información global sobre la satisfacción del alumnado. Si existe validez nomológica, la escala debería mostrar una alta correlación con el indicador 9 de satisfacción global, y además explicar un porcentaje importante de este indicador. Para ello, se ha creado un pequeño modelo estructural que relaciona la escala con este indicador global.

Ilustración 4: Modelo construido para analizar la validez nomológica



Existe validez nomológica si los indicadores de ajuste del modelo son aceptables, para ello se observa en este último paso si el ajuste de este modelo es válido.

En cuanto a la significatividad de las relaciones, el coeficiente *Path* (que analiza el camino o relación entre ambas variables) alcanza un valor de 0,948 (0,869 fue el dato en el estudio anterior). Para contrastar si este valor es estadísticamente significativo se llevó a cabo un proceso de *Bootstrapping* con 5000 submuestras, con cuyos valores se realiza una prueba no paramétrica de significatividad del camino estructural.

El proceso de *Bootstrapping* generó un valor del estadístico *t* para el coeficiente *Path* de 752,61. Este valor es significativo al 99'99% (P-Valor = 0,000). Por tanto, se puede afirmar que sí es significativo y que la relación entre la escala de satisfacción y la pregunta o ítem sobre satisfacción global puede confirmarse estadísticamente.

En segundo lugar, se analiza el coeficiente R^2 . La capacidad del modelo estructural para predecir el comportamiento de las variables endógenas es evaluada por medio de los coeficientes R^2 (Chin, 1998). Este parámetro determina qué cantidad de varianza de los constructos dependientes es explicada por los constructos que la predicen (Chin, 2010). Chin (1998) especifica las fronteras de un poder predictivo bajo, medio y alto de acuerdo a que el coeficiente R^2 alcance los valores de 0'19, 0'33 y 0'67 respectivamente. Si bien en el caso de modelos estructurales que expliquen una variable endógena a partir de pocas variables independientes (por ejemplo, una o dos) un coeficiente R^2 bajo puede ser aceptable (Henseler et al., 2009). En este caso, el coeficiente R^2 alcanza el valor de 0,898 (0,755 en el estudio anterior). Es un valor extraordinariamente alto, la escala de satisfacción explica un 89,8% de la varianza que presenta la pregunta de satisfacción global. Por tanto, se puede afirmar que esta escala tiene una elevada capacidad

de explicación de la satisfacción del alumnado.

Otro aspecto evaluable del modelo estructural está relacionado con su relevancia predictiva, en otras palabras, la capacidad del modelo para predecir el comportamiento de las variables endógenas. La medida predominante para evaluar la relevancia predictiva es el test Q^2 de Stone-Geisser (Stone, 1974; Geisser, 1975) que analiza si el modelo es capaz de predecir adecuadamente los indicadores de los constructos endógenos (Henseler et al., 2009). El valor del estadístico Q^2 se obtiene en PLS a través de un procedimiento denominado *Blindfolding*, técnica de reutilización de la muestra que omite parte de la misma y usa el resultado estimado para predecir la parte omitida (Hair et al., 2011). El test Q^2 da un valor para nuestro caso de 0,829 (0,737 en el estudio anterior).

De manera general, se admite que, si el valor de Q^2 es positivo, el modelo goza de capacidad para predecir el comportamiento de dicha variable (Chin, 2010; Hair et al., 2011). Del mismo modo, mayores indicadores de este parámetro sugieren mejor capacidad predictiva, pudiéndose establecer que los valores 0'02, 0'15 y 0'35 revelan una pequeña, media o elevada relevancia predictiva sobre una determinada variable latente, respectivamente (Henseler et al., 2009). Esto nos lleva a concluir que la escala mantiene una elevada capacidad para predecir la satisfacción del alumnado.

Finalmente, se analiza el coeficiente SRMR como medida de bondad del ajuste del modelo. El valor que alcanza es 0,029. Una evaluación conservadora de este índice considera que se está ante un buen ajuste cuando el índice SRMR se encuentra por debajo de 0,08 (Hu y Bentler, 1999). De acuerdo a ello, el ajuste del modelo alcanza valores de ajuste buenos. El ajuste del modelo con los datos de la encuesta real ha mejorado notablemente. El valor del coeficiente SRMR del estudio exploratorio fue 0,08.

Otros indicadores de ajuste del modelo se presentan en la tabla 13, donde todos los coeficientes de ajusten dan como resultado que el ajuste del modelo es muy bueno.

	Modelo
d_ ULS	0,024
d_ G1	0,052
d_ G2	0,046
Chi-cuadrado	6.441,753
NFI	0,966

Tabla 13: Indicadores de ajuste del modelo nomológico.

Con todo ello, finalmente los valores del modelo causal son muy buenos, y se puede concluir que la escala mantiene validez nomológica para explicar el comportamiento de la variable *Satisfacción del alumnado*.

7 CONCLUSIONES FINALES

Como ya se ha ido indicando a lo largo de este texto, el objetivo de este estudio es presentar unos resultados objetivos acerca de las propiedades psicométricas de la nueva escala de satisfacción del alumnado utilizada en la UEx.

Tras el estudio preliminar, se recomendó la supresión de un indicador para mejorar las propiedades de la escala, lo cual se aceptó, dando lugar a una escala definitiva compuesta por 6 ítems. Esta escala tiene la composición que se muestra a continuación.

- 1 *El docente explica de forma clara y organizada*
- 2 *El docente acompaña adecuadamente las explicaciones teóricas con aspectos prácticos (ejemplos ilustrativos, casos, ejercicios, problemas,...)*
- 3 *El docente usa adecuadamente los recursos didácticos para favorecer el aprendizaje de la asignatura (materiales audiovisuales, campus virtual, pizarra,...)*
- 4 *El docente transmite motivación por el aprendizaje de la asignatura*
- 5 *El docente fomenta la participación activa del estudiante (expresarse en público, reflexionar, defender ideas, trabajar en grupo,...)*
- 6 *El docente resuelve las dudas planteadas y orienta a los estudiantes en el desarrollo de sus tareas*

Los valores de consistencia, fiabilidad y validez de esta escala justifican su empleabilidad, por lo que a partir del Curso 2016/2017 se comenzó a emplear de manera generalizada en las encuestas de satisfacción de la docencia. No obstante, se revisó el estudio una vez se pudo contar con una muestra numerosa tras la oleada de captura de datos del Curso 2016/2017.

Los datos confirman que la encuesta es consistente, presenta una estructura adecuada y es válida para explicar la variable satisfacción en un porcentaje muy elevado. Además, los resultados con esta muestra numerosa han sido mejorados respecto al estudio preliminar, habiéndose obtenido unos valores de fiabilidad, validez y capacidad predictiva de la satisfacción realmente elevados. La conclusión final, por tanto, es que estadísticamente puede afirmarse que el instrumento de medida que utiliza la UEx para medir la satisfacción del alumno con la actividad docente es apropiado y recoge de manera fiable y bastante completa la información sobre esta variable.

Badajoz, diciembre de 2018