



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA

Departamento de Estadística e Investigación Operativa Aplicadas y Calidad

**Propiedades psicométricas de la encuesta de evaluación  
docente e identificación de factores críticos de éxito en  
la satisfacción de los estudiantes con el profesorado**

Autor: Antonio J. Moll Ramis

---

Directores:

Dr. José Miguel Carot

Dra. Andrea Conchado



---

## **Agradecimientos**

Querría dar las gracias en primer lugar a mi familia, por el apoyo recibido, tanto moral como económico, para poder cursar este máster.

A los amigos hechos durante el curso académico, algunos de los cuales formarán parte de mi vida en el futuro.

A mis tutores, José Miguel Carot y Andrea Conchado, por el tiempo que han dedicado a este trabajo, así como por su paciencia en los momentos de zozobra.

## Resumen

En los últimos años hemos asistido a un crecimiento de la preocupación de las universidades por cumplir los estándares de calidad educativa en sus centros, debido tanto al incremento de la exigencia por parte del cuerpo estudiantil como por las leyes aprobadas en los marcos educativos español y europeo. Uno de los indicadores más utilizados para valorar la calidad docente es la encuesta de evaluación de la satisfacción de los alumnos con el profesorado.

El objetivo del presente trabajo es analizar la satisfacción de los estudiantes con la docencia recibida en la Universidad Politécnica de Valencia durante el curso 2016-17. Para ello, se ha validado la escala de satisfacción de la docencia de cuatro factores y un ítem criterio propuesta por el Instituto de Ciencias de la Educación de esta universidad en 2007. Además, se ha confirmado la equivalencia entre el constructo latente medido mediante dicha escala de satisfacción, y un ítem específico sobre el profesor/a evaluado.

Finalmente, se han identificado de forma exploratoria determinados factores influyentes en la satisfacción con la docencia, como la importancia del tipo de asignatura o el nivel de formación del profesor en estudios de Doctorado.

## Resum

En els últims anys hem presenciats un creixement de la preocupació de les universitats per complir els estàndards de qualitat educativa als seus centres, a causa tant de l'increment de l'exigència per part del cos estudiantil com de les lleis aprovades en els marcs educatius espanyol i europeu. Un dels indicadors més utilitzats per a valorar la qualitat docent és l'enquesta d'avaluació de la satisfacció dels alumnes amb el professorat.

L'objectiu del present treball és analitzar la satisfacció dels estudiants amb la docència rebuda en la universitat Politècnica de València durant el curs 2016-2017. Per a fer-ho, s'ha validat l'escala de satisfacció amb la docència de quatre factors i un ítem criteri proposada per l'Institut de Ciències de l'Educació d'aquesta universitat en 2007. A més, s'ha confirmat l'equivalència entre el constructe mesurat mitjançant aquesta escala de satisfacció, i un ítem específic sobre el professor/a avaluat.

Finalment, s'ha identificat de manera exploratòria determinants factors influents en la satisfacció amb la docència, com la importància del tipus d'assignatura o el nivell de formació del professor en estudis de Doctorat.

## **Abstract**

In recent years, we have witnessed an increase in the concern of universities to meet the standards of educational quality in their centers, both because of the increase in the demand on the part of the student body and the laws approved in the Spanish and European educational frameworks. One of the most used indicators to appraise teacher quality is the survey to assess student satisfaction with teachers.

The objective of this paper is to analyse the students' satisfaction with the teaching received at the Polytechnic University of Valencia during the academic year 2016-2017. To this end, the satisfaction scale has four factors and one criterion item, and it was proposed by The Institute of Educational Sciences of the University in 2007. Furthermore, the equivalence between the satisfaction construct, which is measured through this scale, and the criterion item about the professor evaluated was confirmed.

Finally, influential factors in student satisfaction with the education received have been identified in an exploratory way, such as the importance of the type of the subject or the possession of the Doctorate studies.

# Índice

<b>1. Introducción</b>	<b>1</b>
1.1. La medición de la satisfacción con la docencia . . . . .	5
<b>2. Objetivos</b>	<b>9</b>
<b>3. Material y métodos</b>	<b>11</b>
3.1. Instrumentos de medición . . . . .	11
3.1.1. Encuesta de satisfacción de profesorado . . . . .	11
3.1.2. Características del profesor y la asignatura . . . . .	13
3.2. Datos . . . . .	15
3.2.1. Modelo Multinivel . . . . .	18
3.3. Metodología . . . . .	20
3.3.1. Técnicas de validación . . . . .	21
3.3.2. Análisis Factorial Exploratorio . . . . .	21
3.3.3. Análisis Factorial Confirmatorio . . . . .	23
3.3.4. Validez de la escala . . . . .	28
3.3.5. Árboles de decisión . . . . .	29
3.3.6. Software empleado . . . . .	30
<b>4. Resultados</b>	<b>31</b>
4.1. Supuestos estadísticos y análisis descriptivo previo . . . . .	31
4.2. Identificación de la estructura subyacente en las valoraciones de satisfacción del alumnado en el curso 2016-2017 y actividad docente del profesorado . .	34
4.2.1. Análisis de la escala de satisfacción del alumnado con la docencia .	34
4.3. Validación de la escala propuesta por el Instituto de Ciencias de la Educación (2007) en la muestra de valoraciones recopiladas en el curso 2016-2017	36
4.3.1. Fiabilidad y validez . . . . .	47
4.4. Análisis exploratorio sobre los posibles factores que inciden en la satisfacción	50
<b>5. Conclusiones</b>	<b>58</b>

Referencias	60
6. Anexo	65



# 1. Introducción

El mundo universitario ha visto como en los últimos años las ofertas de nuevos grados y másteres han crecido de forma exponencial. Además, la creación de nuevas universidades, tanto privadas como públicas, ha tenido como consecuencia la proliferación de titulaciones. Atrás quedan los tiempos en los que la educación superior en la Comunidad Valenciana se concentraba en la capital, donde esta se impartía en la Universidad de Valencia y la Universidad Politécnica de Valencia, provocando que las familias con más recursos económicos fuesen las únicas que podían acceder a ella. La consolidación de universidades en las provincias de Alicante, con la Universitat d'Alacant y la Miguel Hernández, y Castellón, con la Jaume I, aparte de las universidades privadas, ha llevado a una gran competencia en el sector.

En este contexto, un mayor número de estudiantes pueden acceder a una educación superior cuando acaban la etapa obligatoria de la enseñanza, por lo que las universidades pueden dirigir sus titulaciones a un cuerpo estudiantil con unos antecedentes personales y económicos más heterogéneos que los que se encontraban hace 30 años. Por todo ello, ha crecido la preocupación por mejorar la calidad docente en de los órganos de dirección de las universidades, pues resulta fundamental conocer en qué titulaciones los alumnos perciben una carencia en de calidad en la docencia recibida.

A consecuencia de ello, se ha producido un aumento exponencial de la realización de encuestas, siendo hoy en día habitual que cada alumno tenga la oportunidad de expresar su opinión sobre la calidad de la docencia recibida para cada profesor y asignatura en la que esté matriculado. Debido a este hecho, ha crecido el interés, no solo en crear escalas para evaluar al profesorado, sino que también ha cobrado especial relevancia la validación de dichos cuestionarios, pues los responsables de gestión se preguntan si las escalas utilizadas están, efectivamente, midiendo aquello que pretenden medir.

Ante este contexto social en el sector universitario, el gran objetivo será implementar actuaciones dentro de la universidad con la finalidad de realizar una evaluación de la calidad docente eficiente. Es respecto a este punto donde se produce la primera discusión entre varios de los autores que han investigado sobre este hecho, pues cuando queremos evaluar la labor docente del profesorado universitario nos enfrentamos a la problemática de definir qué es un buen profesor, para en base a ello, ser capaces de conocer cuáles se ajustan a dicha definición y cuáles no.

Desde el punto de vista de Escudero (2006) esto no es el único problema, pues la productividad y motivación de un profesor no es invariante en el tiempo, por lo que cabe desarrollar una metodología de evaluación de estas de forma periódica. Además, sostiene que la productividad en el ámbito docente e investigador son independientes. Por ello, dicha evaluación debe producirse de forma periódica en ambos ámbitos, no resultándonos de utilidad la evaluación investigadora, la cual, debido a las características y motivaciones del mundo universitario, se ha estudiado desde hace más años.

Escudero, Pino y Rodríguez expusieron en 2006, en una línea similar a la planteada por Jornet, que el sistema de evaluación del profesorado debe ser un proceso continuo, con la finalidad de ayudar al profesorado en su desarrollo profesional y en la planificación de su carrera y a contribuir a que la formación en ejercicio se ajuste a las necesidades reales de los profesores. Finalmente, comentar que Martín Moreno (1991) también trabaja en esta línea de actuación.

Una vez se ha establecido que el proceso de evaluación de la calidad debe ser dinámico, cabe definir qué aspectos deben medirse, por ser estos los que más influyen tanto en la satisfacción como en la motivación del estudiante hacia una asignatura en particular. En referencia a este aspecto, Tejedor (2003) considera que un sistema de evaluación del profesorado que pretenda ser válido debe establecer a priori los comportamientos por parte del profesorado que serán deseables, para posteriormente cerciorarse, mediante el proceso de evaluación, si dichas conductas son conseguidas. Algunas de las facetas que se analizan son si el profesor dispone del conocimiento necesario para impartir la asignatura o si este es capaz de relacionarse de forma óptima con el alumnado,. Es decir, no solo es importante que el profesor tenga una buena formación sobre la materia que imparte, sino que además debe ser capaz de expresar sus conocimientos de forma satisfactoria, consiguiendo transmitirlos al alumnado. Además, y ante la posición de algunos investigadores que afirman que la evaluación docente puede llevarse a cabo en función del rendimiento académico de los alumnos, él sostiene que este método es ineficiente.

En referencia a este último punto expuesto por Tejedor, existen casos en la literatura en los que dicha manera de proceder resulta exitosa, aunque son escasos. Por tanto, tendríamos dos posibilidades para medir la calidad educativa. En primer lugar, Eberts, Hollenbeck y Stone (2002) afirmaron que los complementos económicos de carácter individual deberían sustentarse en el rendimiento académico de los estudiantes una vez concluida la asignatura en cuestión. Pese a que esta manera de proceder ha tenido alguna experiencia exitosa (Henry y Rubenstein, 2002), el cuerpo docente es muy reticente a este tipo de evaluación, pues no se ha demostrado una relación directa entre calidad

educativa y resultados académicos. Por otra parte, la segunda opción es sustentada por la gran mayoría de investigadores, entre los que se encuentra Milanowski, quién en 2007 afirmó que la manera más idónea para medir la calidad es mediante factores indirectos. Es decir, se medirá la opinión de los alumnos respecto a un conjunto de variables, sobre las que obtendremos el resultado en los distintos factores.

Un ejemplo de esta manera de proceder sería el trabajo desarrollado por García-Valcárcel (1991, 1992), donde se expone que existen algunas conductas por parte del cuerpo docente que tienen una relación con el rendimiento obtenido por sus alumnos, por tanto, mejorar dichos aspectos resulta de interés. Por ejemplo, comunicar los conocimientos con entusiasmo, establecer canales de comunicación entre ambas partes y tener una planificación de la asignaturas, expuesta previamente al estudiante, clara y concisa aumenta la satisfacción estudiantil significativamente.

Otro aspecto que influye de manera significativa en la satisfacción del estudiante es la corrección de los exámenes, como expusieron Tejedor y García Valcárcel en 1996 y Tejedor en 1998. Según ellos, estos deberían ser corregidos de manera rápida y comentados con los alumnos para que dicha satisfacción no decreciera, siendo percibida una actitud pasiva del profesor nefasta para la satisfacción de este, provocando además una disminución de la motivación hacia la asignatura.

Además, y volviendo a lo expuesto por Tejedor, la motivación del alumnado por aprender crecerá en las asignaturas cuyos profesores tiendan a relacionar conceptos impartidos en el curso con situaciones reales. Esto les permitirá ser capaces de entender la utilidad práctica de los conceptos aprendidos, aumentando el interés por seguir aprendiendo. También resulta un factor significativo en la satisfacción del alumnado la capacidad del profesor para adaptarse al nivel de la clase, es decir, que se adapte la materia que se va a impartir a las capacidades de las que disponen los alumnos. Otro aspecto de interés es el clima de comunicación reinante en la clase, pues si dicha interlocución es constante y fluida, la motivación crecerá. Esto le permite afirmar al autor que existen maneras de proceder del profesorado que implican una mejora en la satisfacción y motivación del estudiante, por lo que la universidad utilizaría la encuesta con la finalidad de detectar una carencia en la consecución de estas facetas, para posteriormente, mediante planes de mejora individualizados y cursos de formación, buscar la mejora de la capacidad docente del profesorado con un déficit en algunos de estos aspectos.

De acuerdo con esta línea de proceder se encuentra Martín Moreno (1991), el cual expone que la universidad debe implicarse de forma activa en el sistema de evaluación y las consecuencias que se deriven de ella, siendo necesario dar un apoyo expreso al cuerpo

docente con la finalidad de que estos perciban el proceso de forma positiva y como una oportunidad de mejorar profesionalmente. Siendo necesario para que esto sea posible que la universidad plantee mejoras en las condiciones de trabajo del profesorado

En línea con lo expuesto por diversos de los autores, Mas (2011) sostiene que existen 6 grandes competencias que debe poseer el profesorado. Entre ellas se encuentran:

- **Diseñar la guía docente:** Este factor engloba todas las actividades que se realizan para elaborar la guía docente, desde diagnosticar las necesidades del alumnado hasta diseñar el plan de evaluación y aprendizaje.
- **Desarrollar el proceso de enseñanza-aprendizaje:** Para satisfacer dicha competencia deben implementarse las TIC (*Tecnologías de la Información y la Comunicación*) para el aula y los trabajos no presenciales. Además, se deben gestionar entornos virtuales y la interacción didáctica.
- **Tutorización:** Engloba el proceso de tutorización de los alumnados para conseguir que mejoren su autonomía y que se fomente un clima de comunicación con el estudiante óptimo. Además, el profesor debe ser accesible en las tutorías de carácter individual.
- **Evaluación del proceso de enseñanza-aprendizaje:** Se incluyen los procesos de evaluación que debe desempeñar el profesorado, así como ser capaz de evaluar la situación del proceso de enseñanza-aprendizaje, lo cual se consigue con la experiencia del profesor.
- **Mejorar la docencia:** Mantener el contacto con el entorno profesional para aplicar situaciones reales a la docencia. Asimismo, se deben implementar técnicas innovadores para mejorar la eficiencia de la docencia.
- **Participación en la dinámica académico-organizativa:** Se cuentan entre sus obligaciones el participar en diversas actividades, como participar en grupos de trabajo y realizar materiales. Además, el docente debe participar en reuniones que tienen como objetivo mejorar la docencia. Para esto, una buena formación tanto docente como pedagógica es necesaria, pues en base a la experiencia se pueden proponer técnicas que puedan ser implementadas por compañeros.

Aunque Escudero, Pino y Rodríguez (2010) coinciden con Martín Moreno en la necesidad de una implicación decidida por parte del cuerpo docente para mejorar la calidad y sostienen que, mientras que una parte del profesorado con problemas de calidad ha hecho uso de estas herramientas para mejorar de forma sustancial su capacidad docente, existen casos en los que estos solo buscan obtener el mínimo necesario para obtener las bonificaciones establecidas.

Con la finalidad de que las universidades mejoraran su calidad, se ha buscado establecer qué posibilidades existen para que el cuerpo docente perciba una recompensa o penalización en función de su actuación. Para ello, Mizala y Romaguera (2004) estudiaron las posibilidades existentes para premiar al profesorado tras obtener los resultados de la evaluación, analizando las ventajas y desventajas de otorgar las bonificaciones de manera individual o colectiva. Por otra parte, Mehrens (1990) expuso varios modelos para evaluar al profesorado, siendo el más extendido en el mundo universitario el modelo de distinción, el cual repercute en un premio o una penalización para el profesor si supera un umbral establecido como mínimo en una serie de variables.

Respecto a las consecuencias que una evaluación negativa tendrían sobre un profesor, la Universidad Politécnica de Valencia recoge en el Manual de Evaluación de la Actividad Docente del Profesorado (Universitat Politècnica de València, 2007) que, de forma general, no se contemplará la suspensión de la docencia de dicho profesor. En base a ello, se proporcionarán, acuerdo con lo establecido por la European Association for Quality Assurance in Higher Education (ENQA), los programas y planes de mejora individualizados que se consideren oportunos a los profesores en cuestión, siendo clave los resultados en las encuestas para sugerir el enfoque de tales planes de mejora.

Pese a todo lo expuesto anteriormente, no solo basta con los planes de mejora implementados por la universidad, sino que es necesario una actitud positiva y de superación personal por parte del profesorado una vez se hayan obtenido los resultados.

### **1.1. La medición de la satisfacción con la docencia**

Una vez se han identificado los factores que influyen de manera más decidida en la satisfacción y motivación estudiantil, se debe proponer un instrumento de medida de dichos conceptos, con el que seremos capaces de identificar comportamientos alejados de los estándares de calidad establecidos por la universidad, así como profesores con grandes capacidades en estos aspectos, los cuales deberían ser reconocidos por parte de la universidad. Aunque hemos visto que el complemento salarial está muy extendido en la literatura, se podría proponer como detalle el envío de una carta personalizada agradeciéndole su buen hacer al docente, pues es un gesto de carácter simbólico que refuerza la autoestima del profesorado.

La mayoría de los autores, entre los que se encuentran Tejedor y Montero (1990), exponen que el mejor método para recoger la opinión del alumnado en relación con la calidad de la enseñanza es mediante el uso de una encuesta. Para ello se deberá diseñar

una escala, donde se recogen diversos ítems relevantes que miden los factores subyacentes deseados, para posteriormente ser validados con la finalidad de conocer si son idóneos para un contexto universitario.

Pese al enorme consenso existente en la literatura sobre la necesidad de usar la encuesta como herramienta principal para evaluar la satisfacción del alumnado con la docencia, existe una cierta divergencia en el número de factores que deberían considerarse. Mientras que Trent y Cohen (1973) consideran que existen 5 factores, centrándose en la motivación, la interacción profesor-alumno y la claridad de las explicaciones y la organización del profesor, Marsh y Overall, en sendos artículos publicados en 1977 y 1980, consideran que existen 7 factores de interés, pues consideran que, además de los arriba mencionados, deben tenerse en cuenta factores que no están relacionados estrictamente con el modo de enseñar.

Por este motivo, resulta interesante comparar la encuesta validada por el Instituto de la Ciencia de Educación de la Universidad Politécnica de Valencia (2007) con otras universidades, lo que realizaremos en el apartado *Instrumentos de medición* de la sección *Objetivos*.

En base a lo anteriormente expuesto, estas divergencias de criterio a la hora de establecer el número de factores que deben formar parte de la escala nos permiten cierta libertad cuando se quieren definir los factores sobre los que los alumnos van a ser preguntados.

Pese a este hecho, existe unanimidad en, una vez establecidos los factores, el procedimiento que se debe seguir para medirlos. Sin duda, las preguntas que componen la escala deben ser claras y concisas, pues en caso contrario podrían llevar a diferentes interpretaciones por los encuestados, aumentando de forma innecesaria el error de medida asociado a cada ítem.

Respecto a este concepto, destacar que los errores de medida se producen por el hecho de que estamos explicando factores subyacentes, los cuales no son medibles directamente, mediante variables observables. Existen varias causas imputables al por qué se dan estos errores de medida. Por una parte, tenemos al investigador que propone la escala, pues de forma subjetiva está seleccionando una serie de ítems con los que, a su parecer, medirá un determinado constructo. Claramente se pueden producir errores de especificación al no considerar algún ítem significativo para medir tal concepto, así como la inclusión en la escala de ítems innecesarios para medir dicho factor, aumentando en ambos casos el error de medida. Por otra parte, cuando un cuestionario es contestado por los encuestados, estos se ven influidos por cuestiones de deseabilidad social, lo que podría deberse a diversos factores. Por ejemplo, si pretendemos medir las causas del alcoholismo entre

los adolescentes, obviar la inclusión en la escala de una pregunta como *Consumo alcohol para mejorar mi imagen social* aumentaría el error de medida a causa del investigador que propone dicha escala. En cambio, personas que no dicen la verdad en el cuestionario pues no quieren reconocer tal hecho como la causa de su alcoholismo, o porque consideran que la encuesta no garantiza lo suficientemente su anonimato y por tanto, declinan contestar acorde a su experiencia,... aumentarían el error de medida, siendo la causa de este incremento imputable al encuestado.

Sin embargo, en las ecuaciones estructurales que vamos a obtener en cada uno de los modelos se diferenciará entre el efecto que cada uno de los ítems tiene sobre el factor con el que está asociado y el error de medida que se asocia a dicho ítem. Por ello, podremos diferenciar entre los ítems que están explicando correctamente el constructo que se pretende medir, de aquellos que están provocando un aumento innecesario del error de medida, por lo que, tras estudiarlos con mayor profundidad podríamos proponer su eliminación de la escala.

Asimismo, pese a que la información que podemos obtener con preguntas del tipo abierto podrían ser de gran ayuda para mejorar la calidad de la docencia, la imposibilidad de otorgarle a estos ítems una escala de medida nos lleva a desechar su utilización en la escala propuesta. Por tanto, las preguntas que se utilizarán serán cerradas y se buscará un grado de objetividad suficiente para que todo el mundo entienda lo mismo al leer una pregunta. Además, estas deben ser relevantes, estando relacionadas con los factores que se pretenden medir, por lo descrito anteriormente respecto al error de medida.

Por otra parte, las preguntas deben ser dotadas de una escala de medida, que nos permita entender la opinión del estudiante y compararla con la opinión del resto de sus compañeros. Para ello, se contará con una de las escalas más sencillas de entender y que cuenta con grandes ventajas, la escala Likert. Esta fue introducida en 1932 (Likert) y presenta, según Llanos, Rosas, Mendoza y Contreras (2001), un sesgo de desviación a la derecha, en especial cuando la población que evalúa está formada por estudiantes. Esto viene provocado, según ellos, por su incapacidad para diferenciar entre las actuaciones del profesor sobre los que son preguntados y su posición hacia dicha materia, por lo que existe una tendencia a puntuar con valores altos y extremos la labor docente. Por otra parte, una de sus grandes ventajas es debido a que es una escala de carácter sumativo, es decir, la puntuación final resultará de la media aritmética de todas las respuestas, lo que nos permite obtener una idea general de la calidad docente del profesor estudiando solamente este valor.

Para poder implementarla, se formularán las preguntas de modo que el encuestado

responda en función del grado en el que está de acuerdo con dicha afirmación, siendo habitual la consideración de 5 posibles respuestas. Además, esta manera de proceder permitirá al encuestado expresar diferencias entre los grados con los que se muestra de acuerdo entre diversas preguntas. En el caso del cuestionario de la Universidad Politécnica de Valencia las posibilidades son las siguientes:

- Totalmente en desacuerdo - 1
- Más bien en desacuerdo - 2
- Ni acuerdo ni desacuerdo - 3
- Más bien de acuerdo - 4
- Totalmente de acuerdo - 5

Finalmente, cabe destacar que, en base a las investigaciones anteriores, queda patente la necesidad de profundizar en el estudio de qué factores están relacionados con la satisfacción, y, por tanto, deben medirse.



## 2. Objetivos

En esta sección vamos a establecer el objetivo principal del trabajo, así como otros de carácter específico. Finalmente, se comentará brevemente la estructura del trabajo.

El objetivo principal de este estudio es **analizar la satisfacción del estudiante con la docencia recibida**. Para cumplir con dicho objetivo, no solo se tendrá en cuenta la escala propuesta por el Instituto de Ciencias de la Educación, sino que además se propondrán varios modelos rivales para realizar una comparación y escoger el modelo que mejor se ajusta a los datos. Para ello es necesario cumplimentar un proceso de conocer con mayor profundidad la estructura subyacente de los datos, validar las escalas con las que contamos y encontrar qué variables influyen significativamente sobre la satisfacción del estudiante. Por todo esto, enunciaremos tres objetivos específicos que deberán ser satisfechos para cumplir con el objetivo principal de este trabajo:

1. En primer lugar, y pese a que los ítems que forman la escala de satisfacción del estudiante están definidos por la teoría, se buscará **explorar la estructura subyacente de los datos, considerando los 8 primeros ítems de la escala**. Estos, según el Instituto de Ciencias de la Educación (2007) fueron formulados para medir 4 factores, siendo el ítem 9 un ítem criterio. Esto nos permitirá conocer con mayor profundidad los datos de los que disponemos, así como realizar comparaciones entre los modelos que son expuestos en la teoría con los modelos que pueden obtenerse en base a este análisis de la estructura subyacente, la cual no tiene por qué coincidir con la teórica.
2. El segundo objetivo sería **confirmar que la estructura de la escala definida por el Instituto de Ciencias de la Educación (2007) de la Universidad Politécnica de Valencia es válida en la base de datos del curso 2016-2017**.
3. Finalmente, el último objetivo, es explorar **qué características influyen de manera significativa sobre la satisfacción del estudiante**. Entre esas características se encuentran tanto las referidas al profesor como las que son propias de la asignatura. Se propone realizar este estudio preliminar con la información disponible y con la técnica de árboles de decisión.

Por último, se explicará brevemente las distintas secciones con las que cuenta el trabajo. A continuación, en el apartado *Material y métodos*, se realizará un análisis descriptivo de los datos para conocer más sobre el perfil de los docentes que son analizados en nuestro estudio y una breve explicación sobre las variables utilizadas. También se hará una comparación sobre el proceso de cambio que se ha producido en la escala de medida de satisfacción del estudiante validadas por el Instituto de Ciencias de la Educación. Además, se explicarán de manera detallada las distintas técnicas utilizadas para cumplimentar con los objetivos enunciados anteriormente.

Por otra parte, en la sección de *Resultados* se mostrarán como las técnicas utilizadas han servido para la consecución de los diferentes objetivos. Asimismo, se propondrán diversos modelos y se producirá a su posterior comparación para seleccionar el que mejor se ajuste a los datos. Cabe recordar que la validación de la escala de medida propuesta por el Instituto de Ciencias de la Educación (2007) para los datos del curso 2016-2017 nos permitiría demostrar la utilidad de esta, así como su validez para distintos años. Por otra parte, analizaremos el efecto que las características del profesor y la asignatura tienen sobre la satisfacción estudiantil.

Finalmente, en el apartado *Conclusiones* se realizará un análisis sobre como se ha desarrollado la consecución de los objetivos y se expondrán las limitaciones e implicaciones del trabajo.

## 3. Material y métodos

En esta sección se realizará un estudio de la evolución de la escala utilizada por la Universidad Politécnica de Valencia para medir la calidad de la docencia impartida y una comparativa con el proceso seguido por la Universidad de Santiago de Compostela. También se explicará el significado de las variables asociadas al profesor y la asignatura sobre las que se realizan las encuestas. Además, se realizará un análisis descriptivo sobre los datos con lo que se trabajará, haciendo hincapié en el perfil del profesorado evaluado. Además, en el apartado *Metodología* se expondrán las técnicas utilizadas para la consecución de los objetivos del trabajo, así como los procedimientos que se siguen en cada uno de los casos.

### 3.1. Instrumentos de medición

#### 3.1.1. Encuesta de satisfacción de profesorado

Antes de entrar con mayor profundidad en la validación de la escala, resulta de interés comentar qué escalas son recomendables para este tipo de situaciones. El Instituto de Ciencias de la Educación (2007) considero que existen 4 factores subyacentes que afectan a la satisfacción. Son los siguientes y aparecen entre paréntesis las abreviaciones que se utilizarán en las figuras que aparecerán a lo largo del trabajo:

- Conocimientos de la materia (Conocimientos)
- Organización y planificación (Organización)
- Metodología docente (Metodología)
- Motivación del profesor (Motivación)

,y consideró que estos factores pueden medirse con 2 ítems cada uno.

Además, el Instituto de Ciencias de la Educación incorpora un último ítem a la escala que funcionará como ítem criterio, como se explicará posteriormente.

En la introducción se han dado a conocer los resultados obtenidos por varios autores, constatando que existe una divergencia tanto en el número de ítems que resultan adecuados para formar la escala de valoración de la satisfacción como en el número de factores que se deben medir. Por ello se va a realizar una comparación entre el proceso de evolución que ha seguido la escala en la Universidad de Santiago de Compostela con el seguido por la de la Universidad Politécnica de Valencia.

En el caso del centro gallego, Tejedor y Montero (1990) publicaron un artículo que recogía, mediante un proceso de investigación bibliográfica, las distintas escalas que se habían considerado

en la universidad. Para medir la satisfacción del estudiante con la docencia, en 1986 se consideró una escala con 49 ítems, los cuales se dividieron en 7 factores, pues eran estos los que obtenían un valor del autovalor superior a la unidad. En el año posterior, el equipo de trabajo encargado de la evaluación del profesorado decidió disminuir a cuarenta los ítems a considerar, lo que provocó una reducción de los factores considerados a 5. En 1989, la estructura de la encuesta había quedado reducida a tan solo 25 ítems, prácticamente la mitad de los que poseía 3 años antes. Se debe tener en cuenta que la realidad social y tecnológica en el contexto universitario ha cambiado de forma notoria, por lo que es recomendable la actualización de dicha escala de forma periódica. Uno de los ejemplos más notorios sería la facilidad actual para proporcionar materiales de calidad a los alumnos, en contraposición con las dificultades existentes hace 30 años, por lo que en la actualidad ítems que midan los materiales proporcionados podrían considerarse con un mayor peso.

En el caso de la Universidad Politécnica de Valencia, como se expone en Martínez (2008), también se ha producido una considerable reducción del número de ítems con los que cuenta la escala. En primer lugar, se tuvo en consideración, en el año 1987, una escala en la que los alumnos habían de responder sobre 34 preguntas, siendo una de ellas un ítem criterio. Cabe destacar que dicho ítem se ha mantenido en el tiempo, y actualmente es el ítem 9 de la escala (*Teniendo en cuenta las limitaciones, pienso que el profesor que imparte esta asignatura debe considerarse un buen profesor*). Tras una primera criba de ítems en 1992, este fue reducido a 19 ítems en 1994, siendo dicha validación del cuestionario llevada a cabo por Martínez (2008). Los sucesivos procesos que las distintas universidades llevaron a cabo para reducir las escalas de las que disponían se deben a varios criterios, entre los que se encuentran la posible confusión del alumnado respecto a una pregunta. Es decir, no se satisfacía el supuesto de que todos los encuestados deben entender lo mismo al leer una pregunta. Otro aspecto sería la falta de peso de un ítem en cualquiera de los factores. Esto puede deberse al proceso de cambio de la sociedad universitaria al largo de los años, que podría llevar a que un aspecto, antes importante, pasara a ser irrelevante para el alumnado.

Actualmente, la escala está compuesta por 9 ítems, la cual fue propuesta por el *Instituto de la Ciencia de la Educación (ICE)* en el año 2007. Seguidamente se mostrarán tanto la redacción actual de los ítems como, entre paréntesis, el título reducido empleado en figuras posteriores del trabajo.

- Parece dominar la materia que imparte (Conocimientos)
- Resuelve dudas y cuestiones sobre la materia con claridad y precisión (Dudas)
- Al inicio del curso proporciona información clara sobre la asignatura: objetivos, programa, metodología y criterios de evaluación (Información)
- Tiene una buena programación del ritmo de las clases y del tiempo dedicado a cada tema

(Programación)

- La metodología empleada y las actividades realizadas en la asignatura ayudan a aprender al alumnado (Metodología)
- Los materiales utilizados y/o recomendados (bibliografía, documentos, recursos didácticos, etc.) son de gran ayuda para el aprendizaje de la asignatura (Materiales)
- Contribuye a crear un buen clima de trabajo y anima a los alumnos a participar en las clases (Clima)
- Su manera de explicar consigue motivar al alumnado y despertar el interés por la asignatura (Motivación)
- Teniendo en cuenta las limitaciones, pienso que el profesor que imparte esta asignatura debe considerarse un buen profesor (Buen profesor)

Cabe reseñar que existe un décimo ítem en el cuestionario, el cual tiene como función medir la calidad de los materiales aportados por el profesor y no fue considerado como parte de la validación de la encuesta. La razón para ello, es que este ítem se considera para evaluar la actividad docente, y no como parte de la estructura de la encuesta. Es decir, la estructura de la escala validada por el Instituto de Ciencias de la Educación solo incluía los 9 ítems enunciados anteriormente. El décimo ítem fue incluido por la universidad para medir otros conceptos, los cuales, en base a los fundamentos teóricos, no están relacionados con el resto. También comentar que, mientras que los ocho primeros ítems son incluidos para medir las estructuras subyacentes consideradas, el ítem 9 supuestamente está midiendo lo mismo que el resto de los 8 ítems. Por tanto, se considera un ítem criterio que nos aporta información sobre si el encuestado está respondiendo de forma coherente con su opinión. Además, se analizará si dicho ítem puede considerarse que esta midiendo al factor subyacente que expresa la satisfacción del estudiante.

### 3.1.2. Características del profesor y la asignatura

Para empezar, se considera interesante contextualizar el proceso de cambio sufrido por las universidades recientemente.

En primer lugar, cabe destacar que durante los últimos años, tanto por el proceso de Bolonia como por la mayor cooperación a nivel educativo entre los países que forman parte de la Unión Europea se ha buscado una unificación de criterio respecto a la calidad de la enseñanza. En el entorno europeo, hemos presenciado la creación del programa DOCENTIA. Este sistema tiene como finalidad crear un marco de competitividad, movilidad y diversidad entre las universidades

europas, buscando mejorar la calidad de manera conjunta y definiéndola con carácter único para el conjunto del mundo universitario europeo.

El primer paso para conseguir una aproximación en los métodos para medir la calidad fue la implantación del programa Bolonia, que se inició en el curso 2010-2011, y que tiene como finalidad unificar titulaciones en toda Europa. Ante este gran paso dado por la comunidad educativa europea a nivel universitario, el programa DOCENTIA adquiere un peso importante por la necesidad de establecer unos parámetros de calidad similares entre los distintos países que forman la Unión.

En este contexto, la Universidad Politécnica de Valencia implementó su Plan Estratégico 2007-2014 (Universitat Politècnica de València, 2007), siendo uno de los objetivos de este la puesta en marcha de mecanismos que permitiesen medir la calidad de la enseñanza, completando al indicador utilizado por excelencia, la encuesta. Además, se presentó la nueva propuesta del Instituto de Ciencias de la Educación que reducía en 10 ítems la escala de 1994. Este plan fue la culminación de un proceso que tuvo lugar tras la aprobación de la Ley Orgánica de Universidades (2001) y su posterior modificación en el congreso (2007), la cual establecía que las universidades están obligadas a realizar una evaluación tanto de la actividad docente como de la actividad investigadora de sus empleados.

En base a esto, se obtuvo que la encuesta de satisfacción no es suficiente para evaluar al profesorado, pues existen otras variables, como podrían ser el grado de madurez o la implicación de estos en su proceso de enseñanza-aprendizaje, que afectan a dicha satisfacción. Por ello, y con las variables que dispone nuestra base de datos, se buscará demostrar que existen características, del profesor y del tipo de asignatura que imparte, que condicionan desde un primer momento la satisfacción con la enseñanza que expresará el alumno al finalizar la asignatura.

A continuación, se presentan las variables que se toman en consideración, así como una breve explicación de ellas. Además, en todas las categorías aparecerá entre paréntesis el título reducido empleado en las figuras que se presentan a lo largo del trabajo. Las variables consideradas son:

1. **Categoría docente:** Se describe el rango del profesor en cuestión, existiendo diversas alternativas. Los niveles de los que consta la variable son:
  - Profesor titular
  - Catedrático
  - Profesor colaborador
  - Profesor ayudante
  - Otros
2. **Tiempo de dedicación:** Se describe si el profesor en cuestión está contratado por la universidad a tiempo completo o parcial.

3. **Doctor:** Se describe si el profesor posee el título de doctorado o no.
4. **Sexo:** Se define si es hombre o mujer.
5. **Sexenios:** Se evalúa el número de sexenios que tiene reconocidos por la universidad.
6. **Años antigüedad:** Variable cuantitativa que expresa el número de años en los que cada profesor ha estado contratado por la universidad.
7. **Curso:** Engloban los tipos de cursos existentes, no existiendo diferenciación entre los cursos de máster y grado. Posteriormente, se verá la necesidad de recodificar los niveles, utilizando del 1 al 5 para los cursos de grado y 6 y 7 para los dos cursos de máster.
8. **Tipo Asignatura:** Esta variable posee varios niveles, entre los que se encuentran si la asignatura sobre la que se evalúa la satisfacción es *Optativa, Básica, Obligatorio o TFG/TFM*
9. **Tipo de Titulación:** Se distingue si el grupo es de *Grado o Máster*, siendo estos los únicos dos niveles que presenta la variable.
10. **Créditos impartidos (Cred Imp):** El número de créditos que el profesor en cuestión ha impartido en una asignatura en particular.
11. **Encuestados:** Variable cuantitativa que describe el número de alumnos que contestaron a la encuesta en cada grupo.
12. **Ítem criterio (*Buen profesor*):** Ítem perteneciente a la escala propuesta por el Instituto de Ciencias de la Educación que mide la satisfacción general.

Dado que uno de nuestros objetivos será buscar qué variables tienen un efecto significativo sobre el ítem *Buen profesor*, utilizaremos el resto de variables anteriormente descritas para identificar qué variables, aplicando un árbol de decisión, provocan que los alumnos posean una mayor o menor satisfacción con la docencia.

## 3.2. Datos

En este apartado vamos a definir el perfil de los profesores evaluados en la base de datos. Por tanto, podremos disponer de mayor información sobre el cuerpo docente de la universidad, sobre los cuales los alumnos expresan su satisfacción con la docencia recibida. Cabe exponer que el alumnado expresa su opinión sobre todas las materias en las que están matriculados, así como todos los profesores que imparten docencia en cada una de las materias. Además, los datos a los que se hace referencia son del curso 2016-2017.

En primer lugar, contamos con 8716 evaluaciones de profesores, habiendo sido algunos de ellos evaluados más de una vez, pues la muestra está formada por 2555 profesores distintos, por

lo que, en media, existen 3.4 evaluaciones por profesor. De estos, 1865 (un 73 %) está empleado a tiempo completo por la universidad. Además, el número de doctores que han sido evaluados es de 2003, representando el 78.4 % del cuerpo docente. El resto, 552 (21.6 %), no disponía del doctorado.

Por otra parte, la experiencia del profesorado es un valor de la Universidad Politécnica de Valencia, pues más de la mitad de sus profesores (50.8 %) han impartido docencia durante 20 o más años, mientras que solamente 244 (9.5 %) tienen 5 o menos años de experiencia docente. Estos datos contrastan con el número de profesores que obtienen el complemento salarial al serles reconocidos los trienios, pues 578 (22.6 %) no tienen reconocido ningún trienio, cuando solamente 244 (9.5 %) tienen una experiencia docente inferior o igual a 5 años. Por otra parte, solamente 29 profesores, lo que representa el 1.1 %, tienen el dato perdido en esta variable. Además, 1254 (49.1 %) tienen reconocidos más de seis trienios, cuando el número de profesores con más de 18 años de experiencia alcanza el 63.6 % (1626 profesores).

Este hecho ocurre de igual manera respecto a los sexenios, donde solamente 461 (18.1 %) tienen reconocidos 3 o más trienios, cuando los profesores con más de 18 años de experiencia, como hemos visto antes, alcanzan el 63.6 %.

Además, el perfil del cuerpo docente es marcadamente masculino, pues hasta 1751 de los 2555 profesores con los que cuenta el estudio son varones, lo que representa un 68.5 % de ellos.

Respecto al número de profesores evaluados según el centro al que están adscritos dentro de la Universidad Politécnica de Valencia obtenemos los siguientes valores:

Centro de adscripción	Frecuencia	Porcentaje
Escuela Técnica Superior de Arquitectura	250	9.785
Escuela Técnica Superior de Ingeniería Agronómica y del Medio Natural	287	11.233
Escuela Técnica Superior de Ingeniería del Diseño	293	11.468
Escuela Técnica Superior de Ingeniería de Edificación	94	3.679
Escuela Técnica Superior de Ingeniería Geodésica, Cartografía y Topografía	46	1.800
Escuela Técnica Superior de Ingeniería Informática	236	9.237
Escuela Técnica Superior de Ingeniería de Caminos, Canales y Puertos	205	8.025
Escuela Técnica Superior de Ingeniería Industriales	383	14.990
Escuela Técnica Superior de Ingeniería de Telecomunicación	117	4.579
Facultad de Administración y Dirección de Empresas	128	5.010
Facultad Bellas Artes	181	7.084
Escuela Politécnica Superior de Alcoi	193	7.554
Escuela Politécnica Superior de Gandía	142	5.558
Total	2555	100

**Tabla 1:** Frecuencias de los profesores evaluados por escuelas



De esta tabla, obtenemos que 2220 (86.89 %) son profesores que desarrollan su actividad docente en el centro que la Universidad Politécnica de Valencia dispone en la capital de la comunidad, mientras que un 7.6 % lo hace en el campus de Alcoi y un 5.6 % en el de Gandía.

A continuación, se presenta una tabla con los rangos de los 2555 profesores analizados, siendo de interés la proporción de asociados (26.73 %) y el número de profesores que disponen de una plaza en la universidad (57.14 %), dato que engloba tanto a los profesores titulares como a los catedráticos.

Descripción	Frecuencia	Porcentaje
Profesor Asociado	683	26.732
Profesor Ayudante	54	2.114
Catedráticos	340	13.307
Profesor Contratado Doctor	265	10.372
Profesor Colaborador	63	2.468
Profesor Titular de Universidad	880	34.442
Profesor Titular de Escuela Universitaria	240	9.393
Otros	30	1.174
Total	2555	100

**Tabla 2:** Frecuencias de los profesores evaluados por categoría docente

Otro aspecto que podría tener influencia en como el alumnado evalúa al profesorado es según el tipo de asignatura o el curso en cuestión. Para ello, presentamos los resultados en la siguiente tabla:

Grupos evaluados	1º curso	2º curso	3º curso	4º curso	5º curso
<b>Grado</b>	1514 17.370 %	1577 18.093 %	1659 19.034 %	1373 15.753 %	112 1.285 %
<b>Máster</b>	2012 23.084 %	411 4.715 %	- -	- -	- -
<b>Idiomas transversales</b>	- -	- -	- -	28 0.321 %	- -
<b>Titulaciones Universidad Politécnica de Valencia</b>	- -	- -	- -	30 0.344 %	- -

**Tabla 3:** Número de grupos que realizan la evaluación del profesor según el curso

Respecto al tipo de asignatura, tenemos que en 1518 casos (17.4 %) se imparte una asignatura del carácter básico, en 4353 esta es obligatoria (49.9 %), en 2790 esta es optativa (32 %) y en 55 casos (0.6 %) se trata de un TFG o TFM.

Por otra parte, tenemos el número de encuestas que han sido contestadas en cada curso, para lo que hay que tener en cuenta que la nota que cada profesor obtiene en un grupo es la media obtenida entre los alumnos de la clase que han contestado al cuestionario.

Personas que participan	1º curso	2º curso	3º curso	4º curso	5º curso
<b>Grado</b>	42086	38768	37743	24329	2125
	22.934 %	21.129 %	20.571 %	13.260 %	1.158 %
<b>Máster</b>	32980	4525	-	-	-
	17.975 %	2.466 %	-	-	-
<b>Idiomas transversales</b>	-	-	-	333	-
	-	-	-	0.185 %	-
<b>Titulaciones Universidad Politécnica de Valencia</b>	-	-	-	591	-
	-	-	-	0.322 %	-

**Tabla 4:** Número de alumnos que realizan la evaluación del profesor según el curso

Con los datos proporcionados en esta tabla podemos realizar una comparación con los datos que se obtuvieron en la tabla anterior. Podríamos sugerir que el número de evaluaciones es superior en los primeros cursos del grado, a pesar de que las diferencias entre el número de grupos que evalúan al profesorado no es excesivamente grande. Pero si lo comparamos con el número de alumnos que responden al cuestionario, podemos afirmar que en los primeros cursos del grado las clases cuentan con un mayor número de alumnos.

Para finalizar, podemos suponer que a medida que se consideran cursos superiores, se requiere una mayor especialización del alumnado, por lo que aumenta el número de asignaturas ofertadas, y por tanto, disminuye el número de evaluaciones por grupo. Otra razón sería que en una misma asignatura, a medida que se avanza en la etapa educativa, es más habitual que más de un profesor se encargue de impartir docencia en ella, por la necesidad de explicar conceptos teóricos variados o por la proliferación de prácticas dentro de una asignatura.

### 3.2.1. Modelo Multinivel

Como hemos visto en el apartado anterior, en la base de datos de las valoraciones de los estudiantes en el curso 2016-2017 disponemos de las evaluaciones de estos de 8716 grupos, siendo

2555 el número de profesores evaluados. Por tanto, nos preguntamos si sería procedente plantear un modelo multinivel para los datos. Consideramos que, por la estructura de la que disponían los datos, los profesores constituyen el nivel más alto, mientras que el conjunto de los grupos que imparte cada profesor forman el nivel más bajo. Con esto, obtenemos que cada grupo está relacionado solamente con el profesor que ha sido evaluado. Además, hemos supuesto que si un profesor tiene una actuación que provoca una alta satisfacción con la calidad de la educación en un grupo, debería obtener una satisfacción similar por parte de los estudiantes en el resto de grupos donde imparte docencia.

La forma de proceder en los modelos multinivel nos permite separar la varianza total de los datos en dos. Por una parte, tenemos la varianza intra grupos, es decir, la varianza que existe entre los elementos que forman parte del mismo grupo, en nuestro caso la variabilidad existente entre los grupos a los que imparte materia un mismo profesor. Por otra parte, tenemos la varianza entregrupos, es decir, la varianza que existe entre los diferentes grupos. En nuestro caso, hacemos referencia a la variabilidad existente entre los grupos en los que imparten materia profesores distintos. Esta partición de la matriz de covarianzas que se produce cuando consideramos este tipo de modelos fue establecida por Bentler (2005). Con ello, disponemos de la matriz de covarianzas intragrupos ( $\sum_W$ ), que nos proporciona las diferencias entre los grupos asociados a un profesor, y la matriz de covarianzas entre grupos,  $\sum_B$ . Y, por tanto, el objetivo será estimar dichas matrices, representando cada una de ellas dos modelos diferentes. Pese a esto, Bentler esgrimió que puede considerarse como un modelo multigrupo que consta de dos niveles, el nivel alto (profesores) y el nivel bajo (notas obtenidas en cada grupo).

Además, como en el caso en que solo consideramos un nivel, Hox (2002) sostuvo que el método más adecuado para realizar la estimación de los parámetros es el de máxima verosimilitud. Por otro lado, Kaplan (1998) y Muthén (1994) exigieron otra condición cuando aplicamos este método, y es que los grupos deben estar balanceados para obtener estimaciones fiables.

El problema principal radica en cuando utilizar estos modelos es adecuado y necesario, para lo que se establece la correlación intraclase, que viene definida como:

$$\text{Correlación intraclase} = \frac{\sigma_B^2}{\sigma_W^2 + \sigma_B^2}$$

Haciendo uso de este índice, Muthén (1997) sostuvo que será adecuado plantear un modelo multinivel cuando exista algún coeficiente intraclase que supere el valor 0.1, pues se intuye que a partir de este valor la existencia de efectos entre grupos es altamente probable. Por otra parte, él mismo consideró que es necesario que cada grupo que consideremos tenga un mínimo de 15 observaciones del nivel más bajo.

A continuación presentamos los índices de correlación intraclase que se han obtenido para los distintos ítems que forman la escala de satisfacción del alumnado con la docencia:

Ítem	Coefficiente de correlación
Conocimientos	0.498
Dudas	0.512
Información	0.434
Programación	0.453
Metodología	0.459
Materiales	0.406
Clima	0.512
Motivación	0.521

**Tabla 5:** *Coefficientes de correlación intraclase de los ítems que forman la escala de satisfacción del alumnado con la docencia*

Pese a que obtenemos índices de correlación intraclase superiores al 0.1 que se plantea en la teoría como umbral para considerar el modelo multinivel, el otro extremo que se considera necesario de que cada uno de los grupos cuente con, al menos, 15 datos, no queda satisfecho.

Cabe destacar que un 40 % de los grupos solo cuentan con un individuo, es decir, existe dos quintas partes de los profesores evaluados (865) que solo han sido evaluados en una asignatura de las que imparten docencia. Además, más de un 32 % de los grupos (693) cuentan con tan solo dos observaciones, obteniendo que tan solo un ínfimo 3 % de los grupos (68) contarían con entre 5 y 8 grupos evaluados, no habiendo ningún docente que cuenta con la evaluación de 15 grupos distintos. Por todo esto, consideramos que, aunque podría tener sentido plantear un modelo multinivel, donde las notas disponibles sean agrupadas según el profesor que las ha obtenido, descartamos esta línea de investigación al no satisfacer de forma clara el mínimo de 15 individuos por grupo exigido en la teoría. Dicha forma de proceder podría ser tenida en cuenta en futuros estudios si tuviésemos la oportunidad de contar con un mayor volumen de datos por profesor, pudiendo ser clave la recopilación de datos de diversos años de los profesores, con lo que podríamos paliar la falta de evaluación de los profesores asociados, pues estos cuentan, en general, con un menor número de créditos impartidos. Finalmente, exponer que los grupos tampoco están balanceados, por lo que de aplicar el modelo multinivel, podríamos obtener estimaciones poco fiables, de acuerdo con Kaplan (1998) y Muthén (1994).

### 3.3. Metodología

En este apartado vamos a profundizar en las distintas técnicas utilizadas a lo largo del trabajo con la finalidad de cumplir los objetivos que nos marcamos. En primer lugar, si aplicásemos el análisis factorial exploratorio y confirmatorio sobre la misma base de datos existiría una tendencia

a que el confirmatorio validase la solución obtenida en el exploratorio, por lo que será necesario la utilización de una técnica de validación. En segundo lugar, necesitamos una técnica que nos permita realizar una reducción de los datos, con la finalidad de conocer con qué estructura subyacente cuentan los datos, por lo que utilizaremos el análisis factorial exploratorio. Por otro lado, para validar los resultados que obtengamos en el exploratorio, haremos uso del análisis factorial confirmatorio. Esta técnica nos permite validar los resultados obtenidos con anterioridad. Además, se explicará como se realiza el proceso de validación de la escala en los distintos análisis realizados. Finalmente, utilizaremos el árbol de decisión para conocer qué características del profesor y la asignatura tienen un efecto significativo sobre la satisfacción del estudiante.

### 3.3.1. Técnicas de validación

En base a lo enunciado por Arruda et.al. (1996), cuando definimos los constructos con un análisis factorial exploratorio y demostramos la validez del constructo aplicando un análisis factorial confirmatorio sobre la misma muestra, de forma general la solución obtenida en el análisis factorial exploratorio será validada por el confirmatorio, lo cual nos provoca un problema que necesitamos solucionar. Por ello, vamos a utilizar la técnica de validación utilizada por Pérez-Gil, Chacón y Moreno en el año 2000. Dichos autores proponen la división de la muestra en dos conjuntos de igual tamaño, es decir, realizan una partición de la muestra en dos submuestras que contienen el 50 % de los datos cada una, utilizando una para realizar un análisis factorial exploratorio y la otra submuestra con la finalidad de confirmar lo establecido en el exploratorio.

### 3.3.2. Análisis Factorial Exploratorio

Como ya hemos establecido en el apartado *Objetivos*, uno de los objetivos de este trabajo es conocer la estructura subyacente existente en los datos de la escala de satisfacción con la docencia. Por tanto, con esta finalidad, realizaremos un análisis factorial exploratorio.

Esta es, según Lévy y Varela (2006), una técnica de reducción de datos que intenta describir un conjunto de  $p$  variables observables mediante un número menor de  $q$  variables no observables. Para este caso se utilizará el primer subconjunto de datos, el cual contiene, como se ha dicho anteriormente, el 50 % de los datos. Aunque uno de los objetivos del trabajo es validar el modelo propuesto por el *Instituto de Ciencias de la Educación*, el análisis factorial exploratorio nos ayudará a entender mejor la estructura de los datos y a proponer otros posibles modelos, pues con él podríamos detectar factores subyacentes diferentes de los establecidos en la teoría.

El hecho de que sea una técnica que se utiliza para reducir la dimensionalidad provoca que, en ciertos ámbitos, se confunda con el Análisis de Componentes Principales. Pero, aunque su objetivo coincida, la manera de proceder es distinta, pues en el caso de aplicar el análisis de componentes principales estamos maximizando la varianza explicada por un conjunto de factores,

que son la combinación lineal de todas las variables, lo que resulta en que la varianza del error sea 0. Esto se contrapone al análisis factorial exploratorio, el cual, mediante un conjunto de factores explica la varianza común de las variables consideradas, y expresa las variables originales como una combinación lineal de los factores considerados. Además, la varianza total de las variables originales será la suma de la varianza común de las variables que son capaces de explicar los factores más el error de medida asociado a cada variable, por lo que solo el análisis factorial exploratorio tiene en cuenta el error de medida, el cual nos será de gran utilidad para proponer reespecificaciones, así como para ser capaces de calcular índices propuestos en la teoría para estudiar la fiabilidad de la escala. En definitiva, la necesidad de conocer el error de medida asociado a cada variable y factor dependiente justifica el uso del modelo de ecuaciones estructurales.

Antes de proceder a analizar los resultados que obtengamos con el análisis factorial exploratorio, existen métodos para conocer si dichos resultados son adecuados o, por el contrario, carecen de sentido y, por tanto, no podemos obtener conclusiones de ellos. Para ello, es necesario que en la matriz de correlaciones puedan observarse agrupaciones entre variables, es decir, existe una correlación entre los distintos ítems que se estudian. Para comprobar si esto ocurre, utilizamos los test de Kaiser Meyer Olikin (KMO) y el test de esfericidad de Bartlett, siendo ambos índices recomendados por Cortina (1993). En el caso del índice KMO, cuánto más próximo sea su valor a 1, mayor correlación existirá entre variables, por lo que el ajuste será mejor. Se considera que si dicho valor no supera el umbral de 0.7 no procede considerar los resultados obtenidos como válidos, siendo 0.9 el valor mínimo que se debe alcanzar para considerar como excelentes los resultados obtenidos. En cambio, el test de esfericidad de Bartlett nos indica que es adecuado proceder con el análisis cuando se rechaza la hipótesis nula, la cual establece que no existe ninguna relación lineal entre las variables que se estudian.

También cabe reseñar que según Floyd y Widaman (1995), las saturaciones en cada factor, una vez hayamos establecido la rotación a considerar, deben superar el umbral del 0.25 cuando en nuestro estudio contamos con más de 300 observaciones, como es nuestro caso, pues disponemos de 4358 grupos en la submuestra que vamos a considerar para aplicar el análisis factorial exploratorio de las variables que forman parte de la escala propuesta por el Instituto de Ciencias de la Educación (2007).

Por otra parte, según Comrey (1988), cuando nuestro objetivo al aplicar el análisis factorial exploratorio sea separar lo máximo posible los factores resultantes, el método de rotación más adecuado será el Varimax. Pese a esto, Nunnally y Bernstein (1995) sostienen que cuando existe una alta correlación entre los distintos factores, suceso que se da en la escala propuesta por el Instituto de Ciencias de la Educación (2007), un método de rotación oblicuo, como son el Promax o Oblimindirecta resultarían más adecuados. Pese a esto, usaremos el método Varimax por creer más conveniente la necesidad de obtener factores lo más separados posibles.

### 3.3.3. Análisis Factorial Confirmatorio

Uno de nuestras finalidades será validar la escala propuesta por el Instituto de Ciencias de la Educación para las evaluaciones de la satisfacción del alumnado en el curso 2016-2017. Para validar dicha escala, se propondrán varios modelos para su validación y comparación. Para realizarlo, nos basaremos tanto en la base teórica (Instituto de Ciencias de la Educación, 2007) como en los resultados obtenidos al estudiar la estructura subyacente de los datos. Para ello, la técnica que vamos a utilizar es el análisis factorial confirmatorio, el cual nos servirá para contrastar que el modelo teórico que el Instituto de Ciencias de la Educación consideró se ajusta a los datos. Las diferencias existentes entre el análisis factorial exploratorio y confirmatorio son plausibles, pues mientras en el primero de ellos todos los factores cargan en cada una de las variables, en el segundo, de forma general, solo un factor carga en una variable. Por esto, en el caso del análisis factorial confirmatorio, el número de factores debe ser establecido de antemano, mientras en el exploratorio se realiza un análisis con carácter exploratorio y se define el número de factores a posteriori, en base a lo congruente que sean los resultados y a como carguen cada uno de los ítems en los distintos factores según el número de componentes consideradas. Además, en el confirmatorio, las relaciones entre los distintos factores en el estudio también son fijadas a priori.

Vamos a proponer modelos de primer y segundo orden. Cuando obtenemos la estructura subyacente de las variables observables, tendremos que estas están agrupadas en factores. Cuando dichos factores están correlacionados entre sí, tenemos que dichos factores están midiendo en la misma dirección. Por ello, se podrá proponer la existencia de un factor que englobe a los factores de primer orden, con lo que cabrá proponer un modelo de segundo orden. En caso contrario, si dichos factores no estuviesen correlacionados, cabría proponer que estos están midiendo actitudes diferentes, por lo que un modelo de primer orden sería adecuado.

Una vez se han explicado de forma breve los modelos de primer y segundo orden, así como las principales diferencias entre los dos métodos mencionados, se explicarán las distintas fases con las que cuenta el análisis factorial confirmatorio.

1. **Especificación del modelo:** Según Hoyle (1995), el objetivo de esta fase es establecer formalmente el modelo que se quiere confirmar posteriormente, para lo que se deben de establecer el número de variables latentes y observables, así como las relaciones existentes entre las diversas variables. Además, dado que las variables no observables carecen de métrica, se recomienda fijar un ítem (normalmente la unidad) para darle una escala de medida a dicho factor (Byrne, 2009).
2. **Identificación del modelo:** En esta fase deberemos comprobar si es posible estimar todos los parámetros que se desean. Hay que tener en cuenta que los parámetros fijados a la unidad para darle métrica a los factores subyacentes no serán tenidos en cuenta en esta

fase.

Para que podamos proceder a la estimación de los parámetros el modelo debe estar sobreidentificado, es decir, el número de grados de libertad debe ser mayor que 0. Para poder realizar esta afirmación, calculamos el número de momentos, que viene dado por  $\frac{p(p+1)}{2}$ , siendo  $p$  el número de variables observables. Para que se cumpla la condición necesaria, que no suficiente, el número de momentos debería ser mayor que el número de parámetros que deseamos estimar. Cabe recordar que en el modelo propuesto se han fijado varios valores, por los que estos no deberían de ser tenidos en cuenta para esta desigualdad.

Por otra parte, según establecieron Long (1983) y Bollen (1989), el cumplimiento de la regla enunciada anteriormente no es suficiente, por lo que se necesita del cumplimiento de otras condiciones suficientes, aunque no necesarias, para poder afirmar que el modelo está sobreidentificado, y por tanto, podemos proceder a la estimación de los parámetros. La primera condición que propusieron fue que cada factor cargue en, al menos, tres ítems. Además, en el caso que existiese correlación entre los distintos factores subyacentes, sería suficiente con que cada factor cargase en dos ítems. Además, la segunda condición sería que cada factor cargue solamente sobre cada una de las variables observables.

Finalmente, para concluir con esta fase, comentar que según Kenny (1979), pese a que el modelo esté sobreidentificado y los parámetros puedan ser estimados, podrían existir estimaciones infractoras que llevarían a estimaciones erróneas de los parámetros. Las más habituales son las siguientes:

- Varianzas de error negativas
- Coeficientes estandarizados que sobrepasan la unidad
- Errores estándar muy elevados.

Por lo que cada vez que obtengamos un modelo que en principio esté sobreidentificado, deberemos comprobar que ninguna de estas situaciones se da en nuestro estudio. El cumplimiento de estas dos circunstancias nos permitirá pasar a analizar los resultados, pues estos serán válidos.

3. **Evaluación del ajuste:** Existen diferentes índices de bondad de ajuste que tienen como finalidad medir cuánto se ajusta el modelo propuesto a los datos. Además, y en función de si se cumple la normalidad multivariante o no, deberemos de hacer uso de los índices calculados mediante el método robusto de Satorra-Bentler  $\chi^2$  de estimación por máxima verosimilitud.

- **Índices de medida del ajuste incremental:** En estos se compara el modelo nulo, es decir, sin la existencia de relaciones entre ninguna de las variables, y el modelo propuesto. Los índices que engloba son:



- **Normal Fit Index, NFI:** Es uno de los índices más ampliamente utilizados. Se considera que adquiere valores aceptables cuando sobrepasa la barrera de 0.9, y que existe un buen ajuste cuando este es mayor que 0.95 (Bentler y Bonet, 1980). En el cálculo se comparan la  $\chi^2$  del modelo teórico con el del independiente.

$$NFI = \frac{\chi_{indp}^2 - \chi_{teor}^2}{\chi_{indp}^2}$$

- **Non Normed Fit Index, NNFI:** Es una variación del NFI, pero teniendo en cuenta los grados de libertad, por lo que este índice es menos sensible que el NFI al tamaño muestral (Turker-Lewis, 1973):

$$NNFI = \frac{\chi_{indp}^2 - \frac{gdl_{indp}}{gdl_{teor}} * \chi_{teor}^2}{\chi_{teor}^2 - gdl_{indp}}$$

- A continuación se presentaran el **Comparative Fit Index, CFI** (Hu y Bentler, 1995), **Bollen's Fit index (IFI)** (Bollen, 1989) y **McDonald's Fit index (MFI)** (McDonald y Marsh, 1990). Estos índices, como los anteriores, oscilan entre 0 y 1 e indican un mejor ajuste del modelo cuanto mayor sean sus valores, pudiendo considerar un buen ajuste cuando se supera el umbral de 0.9, siendo recomendable fijar el umbral en 0.95.

$$\circ CFI = \left| \frac{(\chi_{indp}^2 - gdl_{indp}) - (\chi_{indp}^2 - gdl_{teor})}{\chi_{indp}^2} - gdl_{indp} \right|$$

$$\circ IFI = \frac{\chi_{indp}^2 - \chi_{teor}^2}{\chi_{indp}^2 - gdl_{teor}^2}$$

$$\circ MFI = \exp\left(-\frac{1}{2} * \frac{\chi_{teor}^2 - gdl_{teor}^2}{N}\right)$$

- **Índices de medida de ajuste absoluto:** Tienen como objetivo conocer con qué grado es capaz nuestro modelo de predecir la matriz de varianzas-covarianzas observada. Pese a que en el programa EQS solo dos de estos índices están disponibles, uno de ellos, el *Índice de  $\chi^2$*  resulta extremadamente sensible al tamaño muestral y según James, Mulaik y Brett (1982) este estadístico no es útil cuando tenemos un gran tamaño de muestra, como es nuestro caso. Esta propuesta es corroborada por Jöreskog y Sörbom (1989), los cuales añaden que, ante la falta de normalidad multivariante, para asegurar que el modelo que se propone de ecuaciones estructurales es válido, es recomendable obviar dicho estadístico, tomando en consideración, solamente, los distintos índices de bondad de ajuste propuestos anteriormente. Pese a esto, resulta de gran utilidad cuando nuestro objetivo es simplificar modelos, pues comparando los estadísticos de dichos modelos conoceremos si hemos perdido o no información de manera significativa.

Por otro lado, tenemos el *Root Mean Square Error of Aproximation (RMSEA)*, el cual nos resulta de utilidad. Según Browne y Cudeck (1993), un buen ajuste se obtiene

cuando este indicador es menor que 0.05, aceptable cuando oscila entre 0.06 y 0.08 y mediocre cuando sobrepasa esta barrera. Este índice se calcula de la siguiente manera:

$$RMSEA = \frac{\sqrt{\chi^2 - gdl}}{\sqrt{gdl * (N-1)}}$$

, donde gdl son los grados de libertad del modelo y N el tamaño muestral. Además, destacar que si  $\chi^2$  es menor que los grados de libertad, se establece el valor del RMSEA como cero.

Pese a que podríamos obtener unos altos valores de los índices de bondad de ajuste, podrían existir otros indicadores que nos mostraran que el ajuste realizado no es adecuado. Para cerciorarnos que esto no ocurre, debemos comprobar que los constructos han sido definidos correctamente, por lo que debemos revisar los test de significación para los parámetros estimados, lo que nos permitirá detectar la presencia de algún parámetro en el modelo que no sea significativo. Además, debemos analizar el valor medio de los residuos, así como los índices de fiabilidad que se mostrarán a continuación, los cuales nos serán de gran utilidad para detectar en qué constructos es necesario proceder a una reespecificación.

- El  $R^2$  asociado a cada ecuación, el cual oscila entre 0 y 1 y es mejor cuanto más cercano a 1 esté. Este indicador nos proporciona la parte de la variabilidad asociada a la variable que es explicada por los factores que la afectan, considerándose valores adecuados aquellos que excedan 0.5, según argumentaron Tejedor, Castro y García en 1988.
- Por otra parte, vamos a calcular dos nuevos índices. La fiabilidad compuesta de cada constructo, que nos indicará la consistencia interna de los indicadores asociados a cada factor. Es decir, nos permitirá conocer como de relacionados están entre sí los ítems que forman dicho factor. En este índice, valores por encima de 0.7 son adecuados. Mientras que, para el caso de la varianza media extraída tan solo es necesario obtener valores que superen el umbral de 0.5. En este caso, se nos está indicando que proporción de la varianza total de los ítems es tenida en cuenta por el factor subyacente.

Las formulas para estos dos nuevos indicadores son las siguientes:

- Fiabilidad del constructo p:  $\rho_p = \frac{(\sum_{j=1}^J \lambda_j^p)^2}{(\sum_{j=1}^J \lambda_j^p)^2 + \sum_{j=1}^J \varepsilon_j}$
- Varianza media extraída del constructo p:  $Var_p = \frac{\sum_{j=1}^J (\lambda_j^p)^2}{\sum_{j=1}^J (\lambda_j^p)^2 + \sum_{j=1}^J \varepsilon_j}$

,donde los  $\lambda_j$  son las saturaciones de los ítems en dicho factor y  $\varepsilon_j$  son los errores de medida asociados al indicador en la solución estandarizada.

- Además, para asegurar que en el modelo especificado existe una consistencia interna significativa, es decir, que existe una gran correlación entre los ítems que componen

el modelo, utilizaremos el  $\alpha$  de Cronbach (1951). Este indicador oscila entre 0 y 1 e indica una gran correlación entre las variables cuanto más cercano a la unidad se encuentre. Aunque existen distintos umbrales recomendables para este indicador según el campo de estudio al que nos estemos refiriendo, Nunnally y Bernstein (1995) propusieron 0.7 para nuestro caso.

- Otro indicador que se utilizará para medir la fiabilidad de la escala es el coeficiente ítem-total. Este índice nos permitirá conocer si todos los ítems están midiendo lo mismo. Por otro lado, se podría detectar la existencia de un ítem en la escala cuya contribución a esta fuese no significativa, por lo que cabría proponer su eliminación de la escala.

En los modelos que vamos a especificar prácticamente ningún factor cuenta con grados de libertad suficientes para proporcionar una estimación de sus parámetros de manera individual, por lo que nos resultará imposible conocer el grado de consistencia interna con el que cuentan cada uno de nuestros factores. Por este motivo, mediremos el grado de correlación de las variables para el conjunto de los factores, considerando un modelo de primer orden ,y posteriormente uno de segundo orden.

Por otro lado, destacar que un valor elevado del  $\alpha$  de Cronbach no implica necesariamente la unidimensionalidad de la escala, es decir, los valores elevados que obtendremos, especialmente en los modelos propuestos para la validación de la escala de satisfacción del alumnado con la calidad de la docencia recibida, no implicarán que exista un único factor subyacente que esté englobando a todas las variables observables con las que cuenta la escala. Pese a esto, nos apoyaremos en los índices de bondad de ajuste, los índices de fiabilidad y los test de significación de parámetros para demostrar que, de hecho, en el caso de la satisfacción del estudiante si que existe unidimensionalidad de la escala, y es dicha satisfacción el factor subyacente a todas las variables observables con las que cuenta el modelo.

Finalmente, comentar que, en primer lugar debemos realizar un análisis factorial confirmatorio de primer orden, con el objetivo de validar dicha escala y ver si tiene sentido plantear un modelo de segundo orden, para lo que deberemos estudiar la validez convergente y discriminante. Además, se considerará que en un modelo propuesto no se cumple la normalidad multivariante cuando el estadístico de Mardia (1970) es superior a 5.99. Como este extremo no se cumplirá en ninguno de los modelos propuestos deberemos hacer uso del método robusto de máxima verosimilitud de Satorra-Bentler (1988).

### 3.3.4. Validez de la escala

Uno de los aspectos a tener en cuenta cuando estudiamos una escala de medición es comprobar que existe validez de un instrumento. En este caso no existen un conjunto de índices que nos indican si esta validez de la escala es o no satisfecha, como si ocurría en el caso de la fiabilidad. Por ello, Cronbach y Meelh (1955) expusieron que la validación es un proceso en el que se busca interpretar de manera creíble unos resultados. Para ello, tenemos que demostrar que en nuestra escala existe validez de constructo. Es decir, buscamos cerciorarnos que los factores que componen la escala están midiendo aquello que se pretende medir.

Para ello, tenemos dos tipos de validez que nos permiten asegurar que existe dicha validez del constructo. Por una parte, en base a los conocimientos que poseemos sobre el cuestionario que pretendemos validar, esperaremos que se produzcan una serie de correlaciones entre los distintos factores que han sido establecidos. En este caso, cerciorarnos de la existencia de validez convergente entre dichos factores que esperamos estén relacionados, será un indicativo de validez de constructo. Por otra parte, podemos esperar que existe ausencia de correlación entre dos factores acorde con los conocimientos teóricos, por lo que la comprobación de validez divergente entre los factores en cuestión será un indicador de que existe validez de constructo. Además, esta ausencia de relación entre los factores nos permitirá distinguir entre diversos rasgos.

Tanto la validez discriminante como la convergente fueron propuestos por Campbell y Fiske en 1959. Ellos justificaron la necesidad de comprobar la validez discriminante cuando existe ausencia manifiesta de esta entre los distintos factores. Básicamente, la validez convergente entre constructos se produce cuando existe una relación lo suficientemente elevada como para poder asegurar que están midiendo lo mismo. Esto nos lleva a afirmar que cuando varios constructos están relacionados entre sí, tendría sentido proponer que existe otro factor subyacente que estos constructos estarían midiendo. Por otra parte, si no existe relación entre constructos, podríamos afirmar que están midiendo cosas distintas, luego estos factores estarían discriminados el uno del otro.

Para comprobar la no existencia de validez discriminante se han tenido en cuenta dos métodos, el propuesto por Fornell y Larcker (1981), que es el test de la varianza media extraída frente a las correlaciones. En él, se forma una matriz con los factores y se sitúan las varianzas medias extraídas asociadas a cada factor en la diagonal, y en el resto de posiciones se escriben las correlaciones entre cada par de factores. Para que dicha discriminación exista, la varianza media extraída debe ser el valor más elevado en su columna y fila.

Otro método que se tiene en cuenta es el desarrollado por Anderson y Gerbing (1988) llamado *Intervalo de confianza de las correlaciones*. En él, se halla un intervalo de confianza que, en caso de no contener el 1, asegura que existe validez discriminante entre los factores.

El intervalo es obtenido de la siguiente manera:

$$\text{Corr}(F_i, F_j) \pm t_{0,95}^{\alpha/2} \text{Cov}(\varepsilon)$$

### 3.3.5. Árboles de decisión

Uno de los objetivos específicos de este trabajo es conocer cómo las características de un profesor pueden influir sobre la satisfacción del estudiante, sean variables ajenas o no a su persona. Para cumplir con tal objetivo haremos uso de los árboles de decisión, que es una técnica de minería de datos de aprendizaje supervisado. Es decir, se realizará la división en grupos en base a una variable, en nuestro caso el ítem criterio *Buen profesor*, de la que disponemos previamente. También cabe reseñar que se puede proceder con esta técnica con variables tanto cualitativas como cuantitativas.

Posteriormente, fijaremos las variables dependientes, con las que pretenderemos conocer qué variables influyen significativamente sobre la satisfacción expresada por los alumnos. Todas estas variables han sido definidas en el apartado *Características del profesor y la asignatura* de esta sección.

Esta técnica nos permitirá clasificar en grupos las 8716 evaluaciones de satisfacción del estudiante de las que disponemos en función de las variables dependientes. Este hecho nos permitirá obtener conclusiones sobre cómo diversas variables tienen un efecto sobre la evaluación que realizan los alumnos.

Respecto a esta manera de proceder, Pérez (2011) sostuvo que es una técnica ampliamente utilizada porque facilita la toma de decisiones y explica el comportamiento respecto a una determinada decisión. Esto se debe, principalmente, a que realiza cortes univariantes y el árbol se ramifica en función de la significación de cada una de las variables dependientes, facilitando su interpretación. Con esto, se puede detectar fácilmente cuál es la variable que más influye en la forma de evaluar la satisfacción por parte del estudiante y qué niveles de esta variable se comportan de manera distinta.

El algoritmo utilizado para realizar las particiones en grupos es el Chaid Exhaustivo, el cual se basa en el algoritmo Chaid, propuesto por Kass (1980). Por un lado, tenemos que el método de Kass presentaba algunos inconvenientes, entre los que se encuentran ciertas dificultades para realizar la división óptima. Pese a esto, el Chaid puede generar árboles de decisión eficazmente. Para conseguirlo, se basa en el efecto que tienen las variables independientes sobre la variable dependiente, y las ordena de mayor a menor según el grado de significación que tengan. Posteriormente, realiza la partición en grupos utilizando esa variable, creando grupos excluyentes que nos permiten describir la variable respuesta. De forma similar procede el método propuesto por Biggs (1991), Chaid Exhaustivo, el cual mejora las deficiencias del método de Kass, convirtiéndose en uno de los algoritmos más utilizados actualmente para obtener los árboles de decisión. Antes de la realización del árbol, y para facilitar tanto la interpretación como la obtención de

conclusiones, resulta necesario predefinir la profundidad del árbol. Es decir, cuántas particiones se van a realizar a partir del nodo raíz, siendo este el nodo en el que se agrupan todas las observaciones. En nuestro caso, se consideró que una profundidad del árbol de tres era suficiente. Además, para evitar más particiones de las deseadas se exigirá un mínimo de 30 individuos en el nodo parental. Es decir, si como resultado de una partición tenemos un nodo con menos de 30 observaciones, este no se dividirá en futuras particiones. De igual manera, definiremos un mínimo de 30 individuos en el nodo filial, por lo que si al realizar una partición, uno de los nodos resultantes contiene menos de 30 observaciones, la partición no se producirá.

Finalmente, comentar que el método de validación utilizada para generar los árboles ha sido la validación cruzada. Este procedimiento está basado en subdividir la muestra de manera aleatoria, habiendo sido realizado una partición en 10 grupos de esta. Posteriormente, se construye el árbol con 9 submuestras y utilizamos la que no se utiliza para entrenar el árbol. Tras repetir el proceso en diez ocasiones (no se utiliza la misma submuestra para entrenar el árbol en dos ocasiones), obtenemos el árbol de decisión resultante.

### 3.3.6. Software empleado

Para poder cumplir los objetivos marcados en este trabajo se han hecho uso de diversos programas informáticos, siendo *EQS 6.2* el que más se ha utilizado, pues tanto los distintos análisis factoriales confirmatorios como los cálculos de las relaciones intraclase que se muestran fueron desarrollados con dicho software, proporcionándonos la estimación de parámetros, los índices de bondad de ajuste, las ecuaciones estructurales de cada uno de los modelos,... Además, dado que este software no nos muestra los valores de la fiabilidad del constructo y la varianza extraída del constructo, estos son calculados, a partir de la información proporcionada por el programa *EQS*, con *Excel*.

Por otra parte, diversas particiones se han realizado para proceder con la técnica de validación propuesta por Pérez-Gil, Chacón y Moreno en el año 2000. Para asegurar la aleatoriedad de las muestras hemos utilizado el programa *R*.

También ha sido utilizado el programa *SPSS* para realizar el análisis factorial exploratorio, así como para comprobar los supuestos estadísticos necesarios para proponer un análisis factorial confirmatorio, como son la linealidad de las variables o la correlación entre ellas. Sin embargo, este programa se ha utilizado para satisfacer el objetivo específico 3, en el que buscamos conocer cómo las características de un profesor influyen sobre la satisfacción del estudiante, para lo que se ha utilizado la técnica de los árboles de decisión. Asimismo, las correlaciones ítem-total también han sido calculadas con este software.

Finalmente, para calcular el estadístico F de cada variable en la primera ramificación del árbol de decisión se ha hecho uso del programa *Statgraphics Centurión XVII*.

## 4. Resultados

### 4.1. Supuestos estadísticos y análisis descriptivo previo

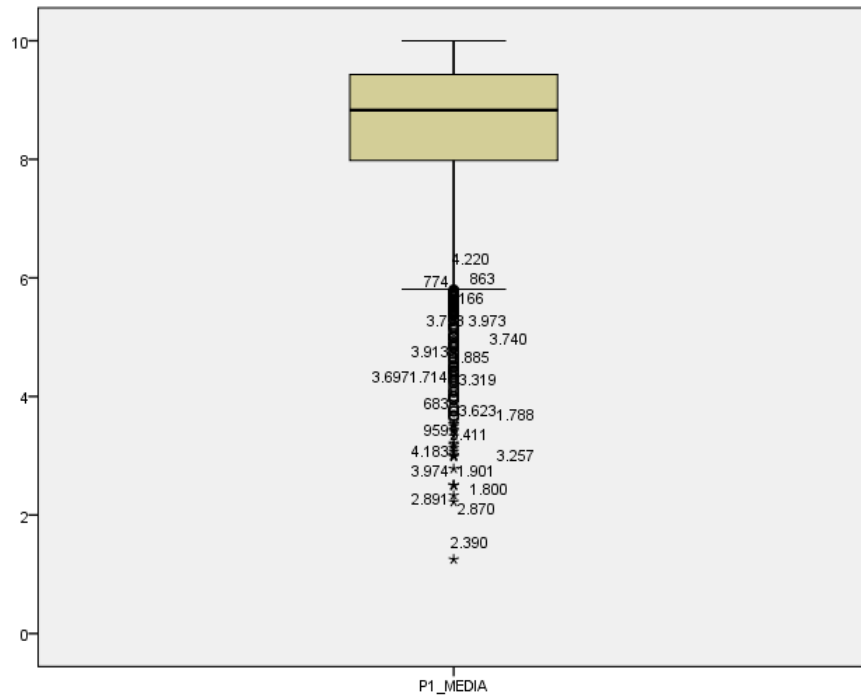
Previamente a los modelos de ecuaciones estructurales que vamos a plantear, cabe comprobar el cumplimiento de varios supuestos estadísticos que son necesarios para obtener estimadores óptimos. Además, se realizará un análisis descriptivo previo de las variables que forman parte de la escala validada por el Instituto de Ciencias de la Educación (2007).

En la siguiente imagen presentamos la media obtenida para cada ítem al considerar todos los grupos que forman parte del estudio, así como su desviación típica.

	N	Mínimo	Máximo	Media	Desviación estándar
<b>Domina materia</b>	4358	1.25	10	8.526	1.239
<b>Resuelve dudas</b>	4358	0	10	7.361	1.698
<b>Información</b>	4358	0.95	10	7.231	1.485
<b>Programación</b>	4358	0	10	7.171	1.601
<b>Metodología</b>	4358	0	10	6.889	1.747
<b>Materiales</b>	4358	0.73	10	6.750	1.531
<b>Clima clase</b>	4358	0.59	10	7.242	1.980
<b>Motivación</b>	4358	0	10	6.665	1.947
<b>Buen profesor</b>	4358	0	10	7.663	1.585

**Tabla 6:** Análisis descriptivo de las variables asociadas a la satisfacción del alumnado con la docencia

Podemos apreciar como, en media, los profesores son aprobados por los alumnos por lo que la satisfacción del alumnado es positiva, siendo el ítem *Parece dominar la materia que imparte* en el que los profesores, de manera general, obtienen una puntuación más elevada. Por otra parte, como veremos posteriormente, dado que ninguna de las variables sigue una distribución normal a nivel univariante, no podemos obtener más conclusiones. Sin embargo, al observar los gráficos box plot, podemos afirmar que en la única variable en la que los profesores obtienen, en más de un 95 % de los casos un aprobado, es la que hace referencia a su dominio sobre la materia que imparte. A continuación, presentamos el gráfico box plot de esta variable.



**Figura 1: Box Plot de la variable 1: Parece dominar la materia que imparte**

El primer supuesto que vamos a estudiar es el de normalidad multivariante. Para ello analizaremos en primer lugar la normalidad univariante de los primeros 8 ítems evaluados por los alumnos. Como el tamaño muestral es mayor a 50 observaciones, el test más adecuado para comprobar dicha normalidad univariante sería el de Kolmogorov-Smirnov, el cual es útil cuando los parámetros de la población son desconocidos, como es el caso. En dicho test, la hipótesis nula es que la distribución que siguen los datos es la misma que la que seguiría una distribución normal. Por tanto, al rechazar  $H_0$  en los 8 casos, tenemos que ninguna de las variables sigue una distribución normal a nivel univariante, por lo que podemos afirmar que tampoco seguirán una distribución normal a nivel multivariante. Este extremo será confirmado en cada análisis que llevemos a cabo por el estadístico propuesto por Mardia (1970), el cual nos indica que los datos siguen una distribución normal multivariante cuando el estadístico es menor que 5.99 con un nivel de significación del 5%. Como en ninguno de los modelos propuestos se satisface dicha restricción, no se dará la normalidad multivariante. Por ello, se tendrá que hacer uso del método de estimación robusto por máxima verosimilitud y los índices de ajuste robustos, entre los que se encuentra el método de Satorra-Bentler (1988).

Otro análisis descriptivo que nos interesa, con la finalidad de conocer si las variables consideradas son indicadores de uno o varios factores, es el estudio de las correlaciones que existen



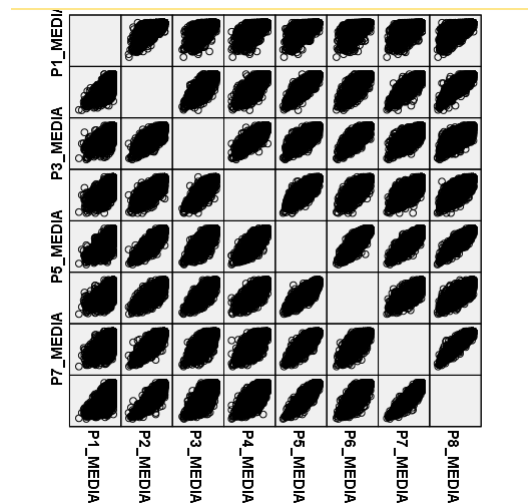
entre cada par de variables. Como podemos ver en la Tabla 7, las correlaciones son bastantes altas entre los distintos pares de variables, y además, en todos los casos son significativas, con un nivel de significación del 1%.

	Domina materia	Resuelve dudas	Proporciona info	Programación	Metodología	Materiales	Clima clase	Motivación	Buen profesor
Domina materia	1	0.890	0.723	0.718	0.728	0.668	0.669	0.727	0.831
Resuelve dudas	0.890	1	0.833	0.835	0.894	0.812	0.878	0.897	0.937
Proporciona info	0.730	0.833	1	0.843	0.816	0.823	0.785	0.783	0.848
Programación	0.718	0.835	0.843	1	0.851	0.787	0.778	0.783	0.858
Metodología	0.728	0.894	0.816	0.851	1	0.871	0.894	0.923	0.923
Materiales	0.668	0.812	0.823	0.787	0.871	1	0.806	0.821	0.842
Clima clase	0.699	0.878	0.785	0.778	0.894	0.806	1	0.942	0.913
Motivación	0.727	0.897	0.783	0.801	0.923	0.821	0.942	1	0.927
Buen profesor	0.831	0.937	0.848	0.858	0.923	0.842	0.913	0.927	1

**Tabla 7:** Correlaciones entre las variables asociadas a la satisfacción del alumnado con la docencia

Otro aspecto que resulta de interés destacar de la tabla es la alta correlación existente entre las distintas variables. Además, debemos destacar que la variable *Dudas* es la que tiene una correlación mayor con *Buen profesor*, que es el ítem criterio establecido por el Instituto de Ciencias de la Educación (2007) para medir la satisfacción general. Por ello, tenemos que la satisfacción del estudiante será elevada cuando, en su opinión, el profesor resuelve dudas de manera clara y precisa.

Otro supuesto estadístico que debemos comprobar es la existencia de linealidad entre cada par de las variables observables involucradas en el modelo. Para ello utilizaremos los diagramas de dispersión de forma matricial.



**Figura 2:** Comprobación de que el supuesto de linealidad se da en nuestros datos

El alto número de observaciones lleva a la superposición de muchas de ellas, pero se puede

apreciar que, de forma general, existe una relación lineal entre todo par de variables.

## 4.2. Identificación de la estructura subyacente en las valoraciones de satisfacción del alumnado en el curso 2016-2017 y actividad docente del profesorado

En este apartado, la finalidad será cumplir el primer objetivo que nos marcamos en este trabajo, es decir, realizar un análisis de los datos para hallar la estructura subyacente de estos. Los datos utilizados serán los que hacen referencia a la satisfacción del alumnado con la docencia en el curso 2016-2017. Para cumplir con esta meta, utilizaremos el análisis factorial exploratorio, una herramienta que nos permitirá conocer con mayor profundidad las relaciones existentes entre las variables y de qué manera estas cargan en los distintos factores.

### 4.2.1. Análisis de la escala de satisfacción del alumnado con la docencia

Como ya hemos dicho anteriormente, nuestro objetivo en este apartado será encontrar la estructura de los datos de la escala de satisfacción del alumnado, para lo que utilizaremos un análisis factorial exploratorio. Pero, antes de obtener conclusiones de los resultados que hemos obtenido, cabe recordar que, según se expuso en el apartado de *Análisis de datos*, dicho análisis puede no ser procedente en ciertos casos, por lo que no tendríamos suficientes fundamentos para obtener conclusiones de los resultados. Con la finalidad de conocer si es procedente utilizar esta herramienta de análisis de los datos, debemos aplicar dos tests. El primero de ellos, la medida de Kaiser-Mayer-Olkin (KMO), adquiere un valor de 0.929, muy cercano a 1, y superior al 0.9 que nos permite afirmar que los resultados obtenidos son excelentes y que existe una gran correlación entre las variables. En segundo lugar, el test de esfericidad de Barlett nos permite, tras rechazar la hipótesis nula de que no existe ninguna relación lineal entre las variables, afirmar que es adecuado proceder con el análisis, lo cual se muestra en la siguiente tabla:

Test	Valor
$\chi^2$	50142.650
Grados de libertad	28
p-valor	0.000

**Tabla 8:** Test de esfericidad de Barlett

A continuación se presentan las comunalidades obtenidas al retener 3 factores y aplicar el método de rotación Varimax. Podemos ver como las comunalidades de las variables son elevadas,

siendo *Conocimientos* la que adquiere un valor menor(0.768). Esto nos indica que la varianza común explicada por la combinación lineal de los factores es elevada, por lo que el porcentaje de varianza asociada a cada variable que no es explicada por los factores considerados es bastante baja.

<b>Variable</b>	<b>Inicial</b>	<b>Extracción</b>
Conocimientos	0.608	0.768
Dudas	0.891	0.925
Información	0.804	0.840
Programación	0.786	0.805
Metodología	0.908	0.919
Materiales	0.794	0.836
Clima	0.900	0.911
Motivación	0.927	0.978

**Tabla 9:** Comunalidades rotadas de las variables de la satisfacción del estudiante cuando extraemos 3 factores

A continuación se presentan los autovalores obtenidos, así como la variabilidad explicada por las componentes considerados.

<b>Factor</b>	<b>Autovalor</b>	<b>Variabilidad explicada</b>
F1	6.695	83.682 %
F2	0.398	4.969 %
F3	0.324	4.052 %
F4	0.212	2.652 %

**Tabla 10:** Autovalores y variabilidad explicada mediante el método de extracción de máxima verosimilitud

Cabría destacar que con solamente un factor seríamos capaz de explicar el 83.7% de la variabilidad, mientras que si nos quedásemos con 3 factores, obtendríamos que estos explican un 5 y un 4.1 % respectivamente. Consideramos que podrían ser seleccionadas las dos siguientes componentes, lo que nos podría llevar a poner en duda el modelo de la teoría, pues en caso de retener 3 factores, los ítems que supuestamente miden los materiales y métodos se dividen.

Ítem	Factor 1	Factor 2	Factor 3
Motivación (P8)	0.805	0.402	0.410
Clima clase (P7)	0.754	0.429	0.400
Metodología (P5)	0.655	0.570	0.408
Materiales (P6)	0.509	0.686	0.326
Información (P3)	0.388	0.670	0.491
Programación (P4)	0.430	0.615	0.491
Domina materia (P1)	0.362	0.352	0.717
Resuelve dudas (P2)	0.569	0.464	0.621

**Tabla 11:** Cargas rotadas de los ítems en el análisis factorial exploratorio con 3 factores

Con todo esto, tendríamos 3 variables latentes:

- Motivación y métodos (Mot-Met)
- Organización y materiales (Org-Mat)
- Conocimientos del profesor (Conocimientos)

Debido a que la estructura subyacente obtenida mediante el análisis factorial exploratorio no coincide con la propuesta con la del Instituto de Ciencias de la Educación, vamos a tomar en consideración ambas formas de proceder en el análisis factorial confirmatorio para, posteriormente, escoger uno de los dos modelos propuestos. Veremos como, efectivamente, el modelo de la teoría es el que mejor se ajusta a los datos, por lo que quedará validado.

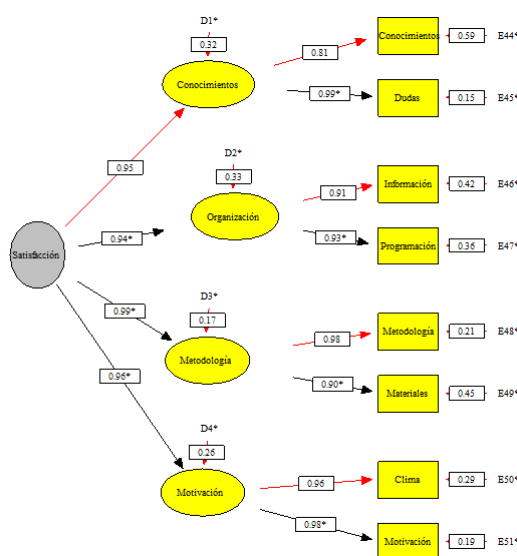
### 4.3. Validación de la escala propuesta por el Instituto de Ciencias de la Educación (2007) en la muestra de valoraciones recopiladas en el curso 2016-2017

En esta sección abordaremos el segundo objetivo del trabajo, que consiste en validar la estructura de la escala que fue definida por el Instituto de Ciencias de la Educación en el año 2007, teniendo en cuenta para ello la satisfacción con la docencia expresada por los alumnos en el curso 2016-2017. Finalmente, se justificará por qué es necesario la consideración de un modelo estructural de segundo orden.

En primer lugar, tenemos que en la propuesta realizada por el Instituto de Ciencias de la Educación (2007) existen 4 factores, compuestos por 2 ítems cada uno, más un ítem criterio (*Teniendo en cuenta las limitaciones, pienso que el profesor que imparte esta asignatura debe*

*considerarse un buen profesor*). Debido a que la estructura subyacente propuesta por el Instituto de Ciencias de la Educación no coincide con la obtenida en el apartado *Análisis de la escala de satisfacción del alumnado con la docencia*, realizaremos una comparación de modelos rivales. Además, se propondrán modelos que incluyan el ítem criterio (*Buen profesor*) para estudiar la idoneidad de incluir dicho ítem en la escala.

Antes de empezar con los distintos modelos que se proponen y comparan, profundizamos en detalle con los pasos con los que cuenta el análisis factorial confirmatorio, los cuales serán obviados en el resto de modelos propuestos. Por simplicidad, se considerará el modelo 3 para realizar este análisis. En dicho modelo, queremos aplicar un análisis factorial confirmatorio de segundo orden, el cual se representa como:

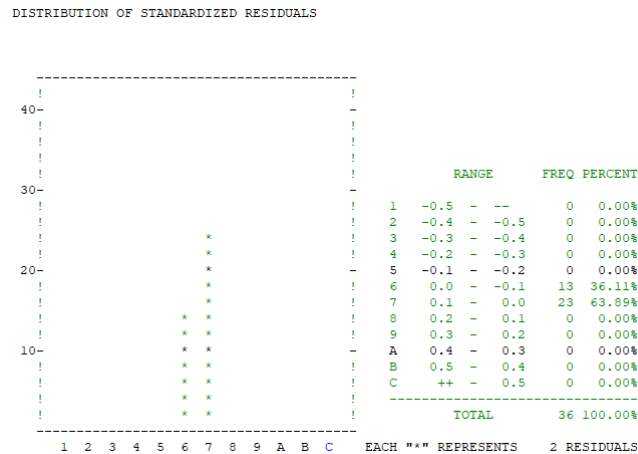


**Figura 3:** Modelo del Análisis Factorial Confirmatorio de 2<sup>o</sup> orden con 4 factores subyacentes de primer orden

En la Figura 3 podemos apreciar como se disponen de 8 variables observables (los ítems que forman parte de la escala), 4 factores subyacentes de primer orden, que representan los conocimientos del profesor, la motivación de este, los materiales y métodos empleados en la asignatura y la programación que sigue. Por otra parte, existe un factor de segundo orden, el cual carga sobre los cuatro factores ya mencionados, siendo su finalidad medir la satisfacción del estudiante. Además, cada factor de primer orden cuenta con un coeficiente de regresión asociado. Cabe destacar que las correlaciones existentes entre los factores que se tienen en cuenta cuando el modelo es de primer orden, desaparecen cuando es de segundo orden. Este hecho se produce porque asumimos que dichas correlaciones son explicadas por el factor de segundo orden considerado. Finalmente, contamos con 8 errores asociados a las variables observables.

Resumiendo, tenemos 8 variables observables, por lo que disponemos de 36 momentos ( $\frac{p(p+1)}{2}$ ) y queremos estimar 4 cargas factoriales que no han sido fijadas a la unidad. Es decir, fijamos un parámetro de cada uno de los cuatro factores a la unidad, para dotar de medida a la escala, y estimamos el otro. Por ello, se estiman las 4 cargas factoriales existentes entre los factores de primer y segundo orden, los 4 errores de regresión, 4 cargas de los factores de primer orden sobre los ítems y los 8 errores de medida. Con todo esto, disponemos de 16 grados de libertad, por lo que al tener un número de grados de libertad positivos y cumplirse las condiciones suficientes establecidas por Long (1983) y Bollen (1989), tenemos que el modelo modelo 3, está sobreidentificado.

Para finalizar con esta breve explicación sobre los pasos seguidos que no mostraremos en el futuro, debemos asegurarnos que, no existen estimaciones infractores, las cuales fueron definidas por Kenny (1979) y, en caso de que no se hallen presentes, podemos afirmar que la estimación de los parámetros es adecuada. Para ello, los residuos estandarizados deben seguir una distribución simétrica centrada en el cero (Figura 4a). También se debe comprobar que todos los ítems, para lo que ponemos como ejemplo el ítem P1, tienen una varianza de su error mayor que 0 (Figura 4b). Y finalmente, debemos comprobar que no existen coeficientes estandarizados que superen la unidad (Figura 4c).



(a) Distribución de los residuos estandarizados

VARIANCES OF INDEPENDENT VARIABLES

-----

STATISTICS SIGNIFICANT AT THE 5% LEVEL ARE MARKED WITH @.

E	D
---	---
E44 -P1_MEDIA	F1
.527*I	.101*I
.012 I	.005 I
43.103@I	18.790@I
( .018)I	( .008)I
( 29.789@I	( 13.339@I

(b) Varianza del error del ítem P1 (Conocimientos)

STANDARDIZED SOLUTION:

	R-SQUARED
P1_MEDIA=V44 = .811 F1 + .586 E44	.657
P2_MEDIA=V45 = .988*F1 + .153 E45	.977
P3_MEDIA=V46 = .908 F2 + .418 E46	.825
P4_MEDIA=V47 = .932*F2 + .363 E47	.869
P5_MEDIA=V48 = .978 F3 + .208 E48	.957
P6_MEDIA=V49 = .895*F3 + .446 E49	.801
P7_MEDIA=V50 = .957 F4 + .291 E50	.915
P8_MEDIA=V51 = .982*F4 + .188 E51	.965
F1 =F1 = .949 F5 + .316 D1	.900
F2 =F2 = .944*F5 + .329 D2	.892
F3 =F3 = .986*F5 + .168 D3	.972
F4 =F4 = .965*F5 + .264 D4	.930

(c) Solución estandarizada: Ningún coeficiente supera la unidad en la solución

**Figura 4:** Comprobación de la no existencia de estimaciones infractoras

Además, dado que el estadístico de Mardia adquiere un valor de 35.439, no existe normalidad multivariante, por lo que la estimación de los parámetros se realizará con el método de máxima verosimilitud de Satorra-Bentler.

Una vez hemos finalizado con la explicación detallada de los pasos que debemos recorrer al realizar un análisis factorial confirmatorio, y antes de realizar la comparación de los distintos

modelos, se calcularán las correlaciones ítem-total. Al realizar estos cálculos cabe comentar que el  $\alpha$  de Cronbach alcanza un valor de 0.977 cuando se consideran los 9 ítems.

Variable	Media de escala si el elemento se ha suprimido	Varianza de escala si el elemento se ha suprimido	Correlación total de elementos corregida	$\alpha$ de Cronbach si el elemento se ha suprimido
Conocimientos	57.039	156.871	0.788	0.978
Dudas	58.190	142.520	0.939	0.972
Información	58.310	149.627	0.870	0.975
Programación	58.388	146.888	0.873	0.975
Metodología	58.658	141.069	0.944	0.972
Materiales	58.785	148.757	0.870	0.975
Clima	58.298	141.222	0.914	0.973
Motivación	58.884	137.070	0.932	0.973
Buen profesor	57.885	143.860	0.968	0.971

**Tabla 12:** Correlaciones ítem-total

Se observa de forma clara como todos los ítems están midiendo lo mismo, pues las correlaciones del ítem con la escala no son inferiores a 0.788. Precisamente, en base a estas correlaciones, podemos afirmar que el ítem *Conocimientos* es el que menos contribuye a la medición de la satisfacción del estudiante, y el ítem *Buen profesor*, el que lo hace de manera más significativa. Además, la eliminación de la variable *Conocimientos* de la escala tendría como consecuencia el aumento de una milésima de la  $\alpha$  de Cronbach, mientras que eliminando cualquier de las otras variables, este índice de consistencia interna disminuiría, lo que ocurre de forma general al reducir el número de ítems de la escala. De esta tabla, el principal resultado es que cuando un profesor obtiene un evaluación alta en el ítem *Buen profesor* por parte del estudiante, la tendencia es que sea evaluado de la misma manera en el resto de la encuesta. Con este hecho, confirmamos la fiabilidad de la escala propuesta del Instituto de Ciencias de la Educación (2007), pues este ítem es el que mejor mide la satisfacción del estudiante con la docencia.

Aunque a continuación propondremos varios modelos para compararlos, nuestro objetivo principal es validar el modelo propuesto por el Instituto de Ciencias de la Educación. Con ello, buscamos validar el modelo de la teoría, y además, demostrar que los otros modelos rivales no son mejores que el propuesto por el Instituto de Ciencias de la Educación.

En primer lugar, consideraremos la posibilidad de que la estructura subyacente esté formada por 1, 3 o 4 factores. Además, en base a lo descrito en el apartado de *Metodología*, plantearemos el modelo de primer orden, para posteriormente y si procede, en base a los resultados obtenidos, plantear un modelo de segundo orden, el cual también deberá ser validado. Además, debido a



que en el marco teórico se definen cuatro factores subyacentes, los resultados del análisis factorial exploratorio indicaban la existencia de tres factores. Por estas razones, buscaremos validar tanto el modelo de la teoría como el resultante de nuestro análisis, concluyendo con una comparación entre dichos modelos y obteniendo cuál se ajusta mejor a los datos.

A continuación, se describen los modelos considerados:

- **Modelo 1:** Modelo de primer orden que considera 4 factores subyacentes, siendo estos medidos por dos ítems cada uno. No se considera la inclusión del ítem criterio.
- **Modelo 2:** Análogo al anterior, pero considerando la existencia de solo 3 factores subyacentes, en línea con lo expuesto al aplicar el análisis factorial exploratorio. No se considera la inclusión del ítem criterio.
- **Modelo 3:** Como veremos con detalle en el apartado *Fiabilidad y validez*, con los resultados obtenidos en el modelo 1, es necesario proponer un modelo de segundo orden. Por ello, tendremos un factor de segundo orden que medirá la satisfacción general del alumnado, sin utilizarse en la medición de este el ítem criterio. También existirán 4 factores de primer orden, medidos con dos ítems cada uno.
- **Modelo 4:** Análogo al modelo 3, pero considerando 3 factores de primer orden en lugar de 4.
- **Modelo 5:** En base a la demostración de que el ítem 9 (*Buen profesor*) es el que mejor mide la satisfacción, se propone un modelo similar al modelo 3, pero con dicho ítem en el lugar del factor de segundo orden.
- **Modelo 6:** Análogo al modelo 5, pero considerando 3 factores de primer orden.
- **Modelo 7 (Propuesta del Instituto de Ciencias de la Educación, 2007):** En este modelo de segundo orden, existe un ítem (*Buen profesor*) que participa en la medición del factor de segundo orden. Además, existen 4 factores subyacentes de primer orden.
- **Modelo 8:** Propuesta realizada por Martínez (2008). Modelo de primer orden con 4 factores subyacentes, en los que cada factor carga sobre dos ítems de manera única. Además, estos 4 constructos cargan a la vez sobre el ítem criterio.
- **Modelo 9:** Análogo al Modelo 7 pero considerando la existencia de tan solo 3 factores subyacentes, en línea con lo obtenido con el análisis factorial exploratorio.
- **Modelo 10:** Como se vio al estudiar la estructura subyacente de los datos, un solo factor explicaba más de un 83 % de la variabilidad, por lo que se considera un modelo de primer orden en el que solo exista un constructo.

- **Modelo 11:** Análogo al modelo 10 pero considerando que el ítem criterio participa en la medición del único factor

A continuación, y en base a lo que se acaba de exponer, se presentan gráficamente los modelos propuestos:

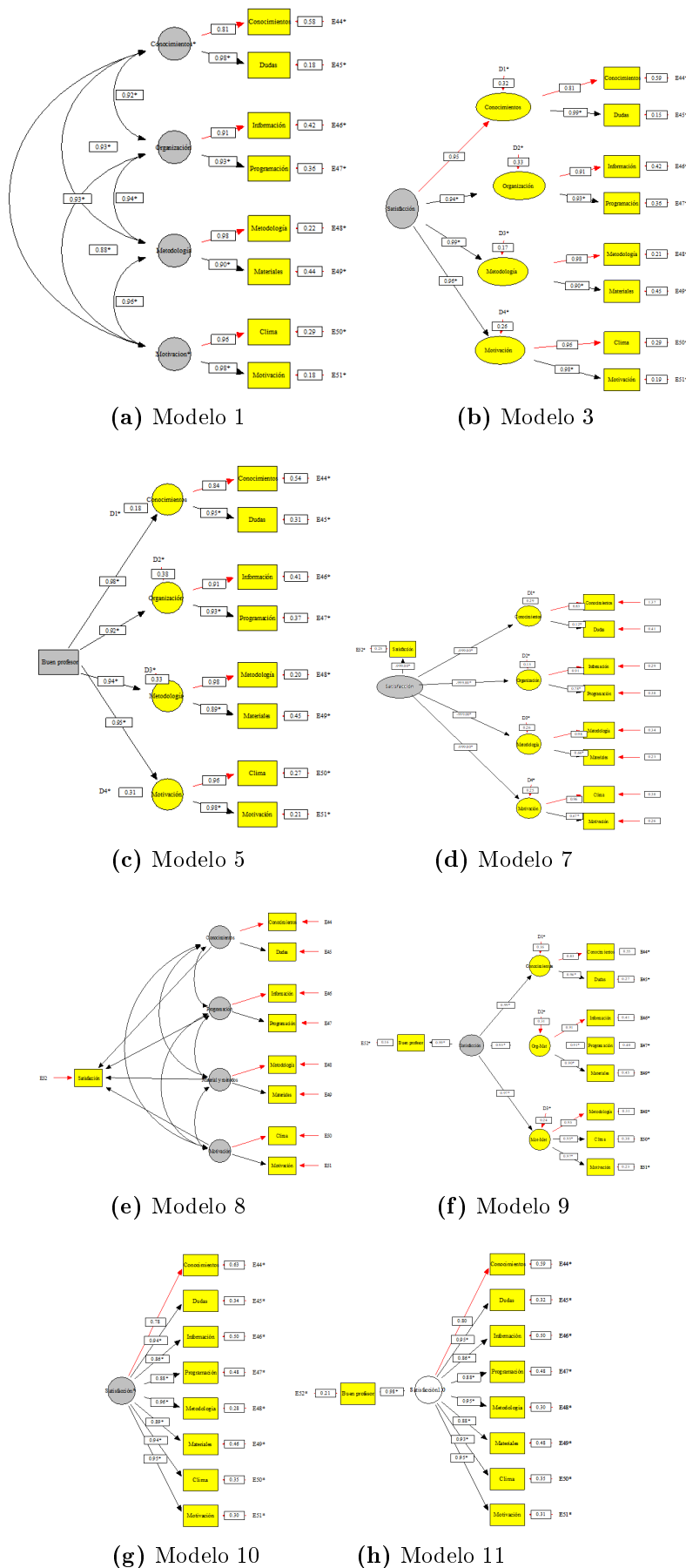


Figura 5: Modelos propuestos para aplicar el análisis factorial confirmatorio

Una vez hemos comprobado que la estimación de los parámetros se ha producido de manera satisfactoria, cumpliendo en todos los casos los criterios pertinentes, tales como la no existencia de estimaciones infractoras, pasamos a comprobar cuánto de buena es la bondad de ajuste en los distintos modelos. Dado la densidad de estos, y con la finalidad de no extendernos en exceso, estos han sido recogidos en la siguiente tabla:

Modelo	NFI	NNFI	CFI	IFI	MFI	RMSEA	$\alpha$ de Cronbach	$\chi^2$ Satorra-Bentler	Grados de libertad	p-valor
1	0.979	0.959	0.979	0.979	0.945	0.090	0.971	507.149	14	0.000
2	0.946	0.913	0.947	0.947	0.865	0.131	0.971	1283.629	17	0.000
3	0.980	0.964	0.980	0.981	0.948	0.084	0.971	478.219	15	0.000
4	0.971	0.950	0.971	0.971	0.924	0.099	0.971	704.121	16	0.000
5	0.931	0.897	0.931	0.931	0.795	0.138	0.977	20121.460	24	0.000
6	0.936	0.909	0.937	0.937	0.810	0.130	0.977	1856.337	25	0.000
7	0.969	0.950	0.970	0.969	0.903	0.096	0.977	912.280	22	0.000
8	0.970	0.942	0.971	0.971	0.907	0.104	0.977	864.380	18	0.000
9	0.977	0.966	0.978	0.978	0.929	0.080	0.977	660.283	23	0.000
10	0.899	0.859	0.900	0.900	0.759	0.166	0.971	2423.928	20	0.000
11	0.898	0.865	0.899	0.899	0.714	0.158	0.977	2969.223	27	0.000

**Tabla 13:** Índices de los distintos modelos considerados

Antes de empezar a comentar los índices de los distintos modelos, cabe recordar que, a excepción del RMSEA, el cual indica un mejor ajuste del modelo especificado a los datos cuanto más pequeño sea, valores por encima de 0.95 nos están indicando que el modelo se ajusta muy bien a los datos. Podemos observar diferencias mínimas en la mayoría de indicadores si comparamos los del modelo 1 (4 factores) con los del modelo 2 (3 factores), pero destaca la caída del índice de McDonald, juntamente con un valor mayor de la RMSEA, lo cual hace inaceptable dicho ajuste (modelo 2) en comparación con el modelo 1. Pues se aleja de manera sustancial del 0.08, valor máximo recomendado para considerar un buen ajuste.

Por otra parte, cabe decir que las altas correlaciones observadas entre los factores en los análisis confirmatorios de primer orden nos lleva a sopesar la existencia de un factor que sea medido por estos factores, lo cual será comprobado en el apartado de *Fiabilidad y Validez*. Por ello, se proponen dos modelos de segundo orden y se observa como el modelo 3, es el que mejores valores obtiene respecto a los índices, por lo que este modelo podría considerarse el más idóneo si consideramos los cuatro primeros y los dos últimos modelos.

Pese a esto, los índices de bondad de ajuste que se alcanzan en el modelo 7 son similares. Tenemos en los tres casos que, los modelos 3, 7 y 9 son los que obtienen el un RMSEA más bajo, pudiendo considerar que se produce un buen ajuste. Hay que recordar que este índice nos indica lo que podríamos esperar si el modelo fuese estimado con la población. Además, hemos demostrado anteriormente que el ítem 9 (*Buen profesor*), no incluido en el modelo 3, es el que

mejor mide la satisfacción del estudiante.

Cabe destacar que la construcción de los índices se basan en la  $\chi^2$ . Además, en los estudios psicométricos los índices CFI,  $\chi^2$  y RMSEA se presentan como principales indicadores, juntamente con el resto. Aunque todos los índices están alineados, en nuestro estudio nos basaremos en el CFI, por su utilidad para detectar diferencias significativas entre modelos de manera inmediata. Esto se debe a que, según Cheung y Rensvold (2002), cuando el índice CFI se diferencia en menos de una centésima, como es el caso, no existen diferencias significativas en este índice. Por tanto, las tres escalas podrían ser consideradas válidas para medir las valoraciones del alumnado respecto a la docencia. Pese a esto, el objetivo principal, como hemos dicho anteriormente, era validar la escala del Instituto de Ciencias de la Educación, hecho que se ha conseguido. Además, se ha demostrado que, aunque existen modelos rivales con índices de bondad de ajuste que superan los umbrales marcados por la teoría, estos no presentan índices mejores que el modelo 7 de manera significativa. Por todo ello, tenemos que la propuesta del marco teórico ha sido validada y que ninguno de los modelos rivales propuestos se ajuste significativamente mejor a los datos.

Además, se observa como, de forma general, los índices de McDonald asociados a cada modelo adquieren el valor más bajo. Asimismo, en aquellos modelos en los que el RMSEA es elevado, los índices de McDonald no alcanzan ni siquiera la frontera del 0.9, por lo que no deberíamos entender que los índices de bondad de ajuste son adecuados. Pese a esto, se podrían entender el caso del modelo 5 y 6, pues se está considerando que existe un ítem que es capaz de medir los 4 factores subyacentes. Por estos motivos, tenemos que los únicos modelos que pueden considerarse válidos son el 3, el 7 y el 9. Esto se debe a que son los únicos que tienen unos índices de bondad de ajuste que superan los umbrales marcados por la teoría. Por ello, se puede considerar que dichos modelos especificados se ajustan a los datos. Por otra parte, y como futura línea de investigación, se podría estudiar la convergencia entre los modelos 3 y 7, pues en base a la similitud de los modelos y los índices obtenidos en ambos podemos creer que sería posible demostrar que el factor subyacente de segundo orden y el ítem 9 convergen.

Finalmente, se realizará una comparación entre las soluciones estandarizadas de dos de los tres modelos arriba mencionados. Aunque también podría haberse considerado el modelo 9, hemos escogido los modelos 3 y 7 por la similitud de la estructura subyacente en ambos casos, dado que solo varía la inclusión en el segundo de ellos del ítem 9 (*Buen profesor*). En primer lugar, tendremos las del modelo 3:

$$\text{Conocimientos} = P_1 = 0,811 * F_1 + 0,586 * E_{44} \quad R^2 = 0,657 \quad (1)$$

$$\text{Dudas} = P_2 = 0,988 * F_1 + 0,153 * E_{45} \quad R^2 = 0,977 \quad (2)$$

$$\text{Informacion} = P_3 = 0,908 * F_2 + 0,418 * E_{46} \quad R^2 = 0,825 \quad (3)$$

$$\text{Programacion} = P_4 = 0,932 * F_2 + 0,363 * E_{47} \quad R^2 = 0,869 \quad (4)$$

$$\text{Metodologia} = P_5 = 0,978 * F_3 + 0,208 * E_{48} \quad R^2 = 0,957 \quad (5)$$

$$\text{Materiales} = P_6 = 0,895 * F_3 + 0,446 * E_{49} \quad R^2 = 0,801 \quad (6)$$

$$\text{Clima} = P_7 = 0,957 * F_4 + 0,291 * E_{50} \quad R^2 = 0,915 \quad (7)$$

$$\text{Motivacion} = P_8 = 0,982 * F_4 + 0,188 * E_{51} \quad R^2 = 0,965 \quad (8)$$

$$F_1 = \text{Conocimientos} = 0,949 * F_5 + 0,315 * D_1 \quad R^2 = 0,901 \quad (9)$$

$$F_2 = \text{Organización} = 0,944 * F_5 + 0,330 * D_2 \quad R^2 = 0,891 \quad (10)$$

$$F_3 = \text{Metodología} = 0,986 * F_5 + 0,167 * D_3 \quad R^2 = 0,972 \quad (11)$$

$$F_4 = \text{Motivación} = 0,964 * F_5 + 0,265 * D_4 \quad R^2 = 0,930 \quad (12)$$

, siendo el factor 5 la *Satisfacción general*

En segundo lugar, se muestran la solución estandarizada del modelo 7:

$$\text{Conocimientos} = P_1 = 0,828 * F_1 + 0,561 * E_{44} \quad R^2 = 0,685 \quad (13)$$

$$\text{Dudas} = P_2 = 0,968 * F_1 + 0,251 * E_{45} \quad R^2 = 0,937 \quad (14)$$

$$\text{Informacion} = P_3 = 0,911 * F_2 + 0,413 * E_{46} \quad R^2 = 0,830 \quad (15)$$

$$\text{Programacion} = P_4 = 0,929 * F_2 + 0,369 * E_{47} \quad R^2 = 0,864 \quad (16)$$

$$\text{Metodologia} = P_5 = 0,979 * F_3 + 0,204 * E_{48} \quad R^2 = 0,958 \quad (17)$$

$$\text{Materiales} = P_6 = 0,894 * F_3 + 0,448 * E_{49} \quad R^2 = 0,800 \quad (18)$$

$$\text{Clima} = P_7 = 0,960 * F_4 + 0,282 * E_{50} \quad R^2 = 0,921 \quad (19)$$

$$\text{Motivacion} = P_8 = 0,979 * F_4 + 0,203 * E_{51} \quad R^2 = 0,959 \quad (20)$$

$$\text{Buen profesor} = P_9 = 0,982 * F_5 + 0,190 * E_{51} \quad R^2 = 0,964 \quad (21)$$

$$F_1 = \text{Conocimientos} = 0,981 * F_5 + 0,193 * D_1 \quad R^2 = 0,963 \quad (22)$$

$$F_2 = \text{Organización} = 0,944 * F_5 + 0,330 * D_2 \quad R^2 = 0,891 \quad (23)$$

$$F_3 = \text{Metodología} = 0,969 * F_5 + 0,243 * D_3 \quad R^2 = 0,938 \quad (24)$$

$$F_4 = \text{Motivación} = 0,965 * F_5 + 0,261 * D_4 \quad R^2 = 0,932 \quad (25)$$

Si comparamos las ecuaciones de ambos modelos, se puede observar que al incluir el ítem *Buen profesor* para medir el segundo factor, que describe la satisfacción general, se producen variaciones muy pequeñas en los coeficientes. Pese a esto, si que existe una pequeña variación en los coeficientes de dos de los factores de primer orden. Por una parte, tenemos que en el modelo 7, el factor *Conocimientos* se mide con mayor precisión. En este caso, tenemos un error de regresión inferior, y un coeficiente de determinación que asciende de 0.901 a 0.963. Por otra parte, en el modelo 3, el factor *Satisfacción general* explica un porcentaje de variabilidad (97.2 % frente a 93.2%) superior en el caso del factor *Conocimientos*, por lo que tenemos que cada

modelo mide con mayor precisión uno de los factores y no se hallan diferencias significativas en los coeficientes de los otros dos. Todos estos resultados confirman lo expuesto anteriormente, pues ambos modelos miden de manera excelente la *Satisfacción general*, con pequeñas diferencias. Además, es destacable el error de medición asociado al ítem *Buen profesor*, pues es el más pequeño de las variables observables. Esto nos permite afirmar que la redacción de este ítem criterio es la adecuada, por lo que no se requiere ninguna reformulación de este.

Finalmente, y en base a lo expuesto en este apartado, concluimos que la propuesta teórica del Instituto de Ciencias de la Educación (2007) ha sido validada y presenta índices de bondad de ajuste que superan los umbrales establecidos por el marco teórico. Además, ninguno de los modelos rivales consigue unos índices de bondad de ajuste que resulten significativamente mejores que los obtenidos por el modelo teórico.

#### 4.3.1. Fiabilidad y validez

En este apartado nos vamos a centrar en los modelos validados en el apartado anterior. Para ello, se calcularán los índices de fiabilidad enunciados en *Metodología* ( $R^2$ , varianza media extraída y fiabilidad del constructo) para el modelo 7. Además, se presentará la justificación de la necesidad de considerar un modelo de segundo orden (Modelo 3) en base a los resultados que se obtuvieron en el modelo de 4 factores y de primer orden (Modelo 1).

En primer lugar, cabe recordar que existen varias maneras de medir la fiabilidad de cada constructo. Es decir, que los ítems están midiendo dicho factor y que además presentan una alta correlación entre ellos. Para demostrar la existencia de fiabilidad calculamos anteriormente las correlaciones ítem-total, que nos permitieron afirmar que el ítem *Buen profesor* es el que mayor correlación tiene con la escala. Por tanto, su contribución a la medición de la satisfacción es la más significativa. En este apartado, calcularemos otras medidas de fiabilidad.

Uno de los índices utilizados es el coeficiente de determinación asociado a cada ecuación factorial, en el se mide la proporción de varianza del ítem que es explicada por el constructo que lo contiene. Para estudiar estos tres índices debemos tener en cuenta las soluciones estandarizadas del modelo 3 y 7. Aunque en general obtenemos que los ítems son buenos indicadores de las variables latentes a las que están asociadas, cabe destacar que el primer ítem, que evalúa si el profesor conoce la materia que imparte, es solamente explicado en un 66 % por su variable latente asociada en el modelo 3. Este porcentaje se eleva hasta el 69 % en el caso del modelo 7, lo que corrobora lo observado anteriormente de que el factor *Conocimientos* se explica mejor en este modelo. Pese a esta pequeña puntualización, los altos valores de las  $R^2$  asociadas a las distintas ecuaciones, nos indican que una gran proporción de su variabilidad está siendo explicada por su factor asociado, por lo que no procede a proponerse ninguna reespecificación del modelo. Además, el valor del  $\alpha$  de Cronbach se calculó en el apartado anterior, y era adecuado. Por ello, podemos confirmar la existencia de consistencia interna entre los ítems de la escala.

Seguidamente, vamos a realizar los cálculos necesarios para obtener el valor de los índices de fiabilidad y la varianza media extraída del constructo. Los resultados que hemos obtenido para estos dos indicadores vienen dados al desarrollar el modelo 7. Es decir, un análisis factorial confirmatorio de segundo orden considerando 4 factores subyacentes con un ítem criterio participando en la medición del factor de segundo orden, que son:

<b>Factor</b>	<b>Fiabilidad del constructo</b>	<b>Varianza media extraída</b>
F1	81.360 %	68.708 %
F2	81.256 %	68.431 %
F3	84.287 %	72.882 %
F4	88.699 %	79.695 %
F5	94.839 %	78.614 %

**Tabla 14:** *Fiabilidad de los constructos y varianza media extraída por cada uno de los factores en el Modelo 7*

Se obtiene que todos los factores superan el umbral del 0.8 establecido para la fiabilidad del constructo. Por ello, podemos afirmar que los ítems que forman cada uno de los factores están correlacionados entre sí. Cabe recordar que, según Nunnally y Bernstein (1995), el umbral mínimo es 0.7. Además, cabe destacar el valor obtenido para el factor 5 (que mide la satisfacción del estudiante) lo cual nos indica que resulta muy adecuado realizar un análisis factorial confirmatorio de segundo orden con la inclusión del ítem 9, pues los factores de primer orden tienen un gran relación entre sí, y esto da lugar a la creación de otro factor subyacente que es medido con estos 4 factores. Por otra parte, el valor de varianzas medias extraídas es reseñable en todos los casos, pues la menor puntuación las obtienen los factores 1 y 2 con 0.68, bastante superior al 0.5 (valor mínimo recomendable). Por tanto se puede afirmar que una gran parte de la variabilidad total de los ítems es tenida en cuenta por las estructuras subyacentes a los que están asociados, siendo los factores 1 y 2 los que menos variabilidad explican, y el factor 4 el que más. En el caso del factor 1, este hecho puede ser imputado al error de medida asociado al indicador 1, que pese a no ser excesivamente elevado, si lo es en comparación con el resto de errores que tienen el resto de ecuaciones.

Por otra parte, queremos demostrar que era necesario la realización de un modelo con 4 factores de segundo orden. Para ello, utilizaremos los métodos descritos en *Metodología*, por lo que deberemos de calcular la fiabilidad del constructo y la varianza media extraída para la solución estandarizada del modelo 1.



<b>Factor</b>	<b>Fiabilidad del constructo</b>	<b>Varianza media extraída</b>
F1	81.411 %	68.857 %
F2	81.256 %	68.433 %
F3	84.287 %	72.880 %
F4	88.699 %	79.696 %
F5	93.203 %	77.422 %

**Tabla 15:** *Fiabilidad de los constructos y varianza media extraída por cada uno de los factores en el Modelo 3*

Tras obtener el valor de los índices de fiabilidad para el modelo 3, utilizamos dichos valores para realizar los test que nos permitirán asegurar que existe validez convergente, y por tanto, es necesario plantear un modelo de segundo orden.

Con el primero de ellos, se comprueba que la varianza media extraída no es el valor más elevado de su fila y columna correspondiente, por lo que no existe validez discriminante entre los factores. Este resultado es consistente con la validez del modelo 3 con un constructo de segundo orden, que relaciona los cuatros factores anteriores. Por todo ello, y en base a los resultados obtenidos en el modelo 1, confirmaremos la necesidad de plantear un modelo de segundo orden.

	<b>F1</b>	<b>F2</b>	<b>F3</b>	<b>F4</b>
<b>F1</b>	0.689	0.922	0.927	0.926
<b>F2</b>		0.684	0.941	0.880
<b>F3</b>			0.729	0.957
<b>F4</b>				0.797

**Tabla 16:** *Matriz para comprobar la existencia de validez discriminante según el método de Fornell y Larcker (1981)*

El otro método utilizado se basa en calcular los intervalos de confianza de las correlaciones entre constructos. En nuestro caso, podemos observar como solo en dos intervalos de confianza entre los distintos pares de factores el uno no está contenido, y no lo está por muy poca diferencia, por lo que podríamos confirmar nuevamente que no existe validez discriminante entre los factores y se puede proponer de forma razonada un análisis factorial confirmatorio de segundo orden en base a dichos factores están relacionados y están midiendo otro factor subyacente.

Parejas de factores	Correlación	S.d.Covarianza	Mínimo	Máximo
<b>F1-F2</b>	0.922	0.033	0.857	0.987
<b>F1-F3</b>	0.927	0.040	0.849	1.010
<b>F1-F4</b>	0.926	0.041	0.846	1.010
<b>F2-F3</b>	0.941	0.051	0.841	1.041
<b>F2-F4</b>	0.880	0.050	0.782	0.978
<b>F3-F4</b>	0.957	0.064	0.832	1.082

**Tabla 17:** Matriz para comprobar la existencia de validez discriminante según el método de Anderson y Gerbing (1988)

#### 4.4. Análisis exploratorio sobre los posibles factores que inciden en la satisfacción

En este apartado, tenemos como objetivo realizar un análisis exploratorio sobre cómo características del profesor y de la asignatura en la que se realiza la evaluación de la satisfacción tienen una influencia sobre esta. Para cumplir con dicho objetivo, se utilizará la técnica de minería de datos del árbol de decisión.

En primer lugar, antes de realizar el árbol de decisión, realizaremos un estudio previo de las variables de las que disponemos. La finalidad es, en base a tablas de contingencia, estudiar los posibles efectos que se encuentran confundidos en las variables. Claramente, la variable que mayor atención nos ha suscitado es *Curso*, pues no existe una diferenciación entre las evaluaciones de los grupos de primer curso de Grado y Máster. Para comprobar si dicho efecto está de hecho confundido presentamos la siguiente tabla:

Curso	Grado	Posgrado	Total
1	1544	2012	3556
% en curso	43.420 %	56.580 %	100 %
% en titulación	24.535 %	83.038 %	40.799 %
2	1577	411	1988
% en curso	79.326 %	20.674 %	100 %
% en titulación	25.096 %	16.962 %	22.809 %
3	1659	0	1659
% en curso	100 %	0 %	100 %
% en titulación	26.363 %	0 %	19.034 %
4	1401	0	1401
% en curso	100 %	0 %	100 %
% en titulación	22.263 %	0 %	16.074 %
5	112	0	112
% en curso	100 %	0 %	100 %
% en titulación	1.780 %	0 %	1.285 %
Total	6293	2423	8716
% en curso	72.201 %	27.799 %	100 %
% en titulación	100 %	100 %	100 %

**Tabla 18:** *Tabla de contingencia de las variables Curso y Tipo de titulación*

Se observa claramente como se cumple lo que preveíamos. De las 3556 evaluaciones de grupos del primer año en la titulación correspondiente, un 56.6 % son de máster, mientras que el 43.4 % pertenecen a grado. Este hecho puede tener una gran confusión pues las características de los estudiantes de primero de máster y grado pueden ser muy diferentes. Por ello, se recodificarán los grupos de primero de máster con el nivel 6. Por la misma razón, se diferenciará entre los grupos de segundo curso de grado y máster. Por otro lado, existen 112 evaluaciones del quinto curso de grado. Esta situación es muy particular pues solamente existen dos titulaciones que requieren dicho curso, que son *Grado en Arquitectura* y *Grado en Fundamentos de la Arquitectura*. Aunque podría considerarse su eliminación de la muestra en base a las grandes diferencias del tamaño del grupo respecto al resto de opciones, se considera que es lo suficientemente grande como para mantenerlo en el trabajo. Esto se debe principalmente a que los tamaños considerados como parental y filial, como se ha expuesto en *Metodología* son muy inferiores al tamaño de dicho grupo.

Tras realizar esta primera recodificación, analizamos el peso que tienen los distintos tipos de Asignatura (*Básica, Obligatoria, Optativa o TFG/TFM*) en cada curso. Procediendo de igual

manera, obtenemos la tabla de contingencia siguiente:

Curso	Básica	Obligatoria	Optativa	TFM	Total
1	1273	187	84	0	1544
% en curso	82.448 %	12.111 %	5.440 %	0 %	100 %
% en tipo titulación	83.860 %	4.296 %	3.0 %	0 %	17.725 %
2	231	1275	71	0	1577
% en curso	14.648 %	80.850 %	4.502 %	0 %	100 %
% en tipo titulación	15.217 %	29.290 %	2.545 %	0 %	18.104 %
3	14	1199	446	0	1659
% en curso	0.844 %	72.272 %	26.884 %	0 %	100 %
% en tipo titulación	0.922 %	27.544 %	15.986 %	0 %	19.045 %
4	0	461	935	5	1401
% en curso	0 %	32.905 %	66.738 %	0.357 %	100 %
% en tipo titulación	0 %	10.590 %	33.513 %	9.091 %	16.083 %
5	0	76	36	0	112
% en curso	0 %	67.857 %	32.143 %	0 %	100 %
% en tipo titulación	0 %	1.746 %	1.290 %	0 %	1.286 %
6	0	1093	869	50	2012
% en curso	0 %	54.324 %	43.191 %	2.485 %	100 %
% en tipo titulación	0 %	25.109 %	31.147 %	90.909 %	23.097 %
7	0	62	349	0	411
% en curso	0 %	15.085 %	84.915 %	0 %	100 %
% en tipo titulación	0 %	1.424 %	12.509 %	0 %	4.718 %
Total	1518	4353	2790	55	8716
% en curso	17.416 %	49.943 %	32.010 %	0.631 %	100 %
% en tipo titulación	100 %	100 %	100 %	100 %	100 %

**Tabla 19:** Tabla de contingencia de las variables *Curso* y *Tipo de asignatura*

En esta tabla podemos observar como existen efectos confundidos pues tenemos que las optativas están concentradas, de forma mayoritaria, en los cursos de tercero y cuarto de grado y en los dos cursos de máster. Por otro lado, el 83.9 % de las asignaturas *Básicas* se imparten en el primer curso de grado, por lo que la satisfacción del estudiante podría depender de varios factores. De igual forma, las asignaturas del tipo *Obligatorias* se concentran en el segundo y tercer curso de grado, y el primero de máster. Finalmente, se eliminarán las evaluaciones de *TFG*, pues el número de evaluaciones (5) es inferior al establecido para los nodos parental y filial, y por

tanto, no podría producirse ninguna ramificación.

Con estas reformulaciones de los niveles de las variables, se considera que los efectos confundidos han sido reducidos significativamente. Pese a ello, debe quedar constancia de su existencia, siendo la concentración del tipo de asignatura en algunos cursos una de las mayores confusiones. Por este motivo, las interpretaciones que se realicen del árbol deben estar basadas en los conocimientos de los expertos, y deben producirse con un conocimiento exhaustivo del proceso enseñanza-aprendizaje.

A continuación, se presenta el árbol de decisión obtenido al considerar como variable dependiente el ítem 9 (*Buen profesor*), el cual forma la escala propuesta por el Instituto de Ciencias de la Educación (2007). Para realizar la elección de esta variable, nos basamos en los resultados obtenidos anteriormente, pues quedó validada la propuesta del Instituto de Ciencias de la Educación. Con ello, se demostró que el ítem 9 es un buen indicador de la satisfacción general, por lo que será utilizado. Aunque se podrían considerar otras variables, habría que considerar que en el análisis factorial los pesos de estas dependerían de la variabilidad de cada factor y de las cargas que cada variable de la escala tendría en dichos factores. Cabe destacar que, debido a la gran dimensión del árbol, se presenta en dos figuras para facilitar su interpretación.

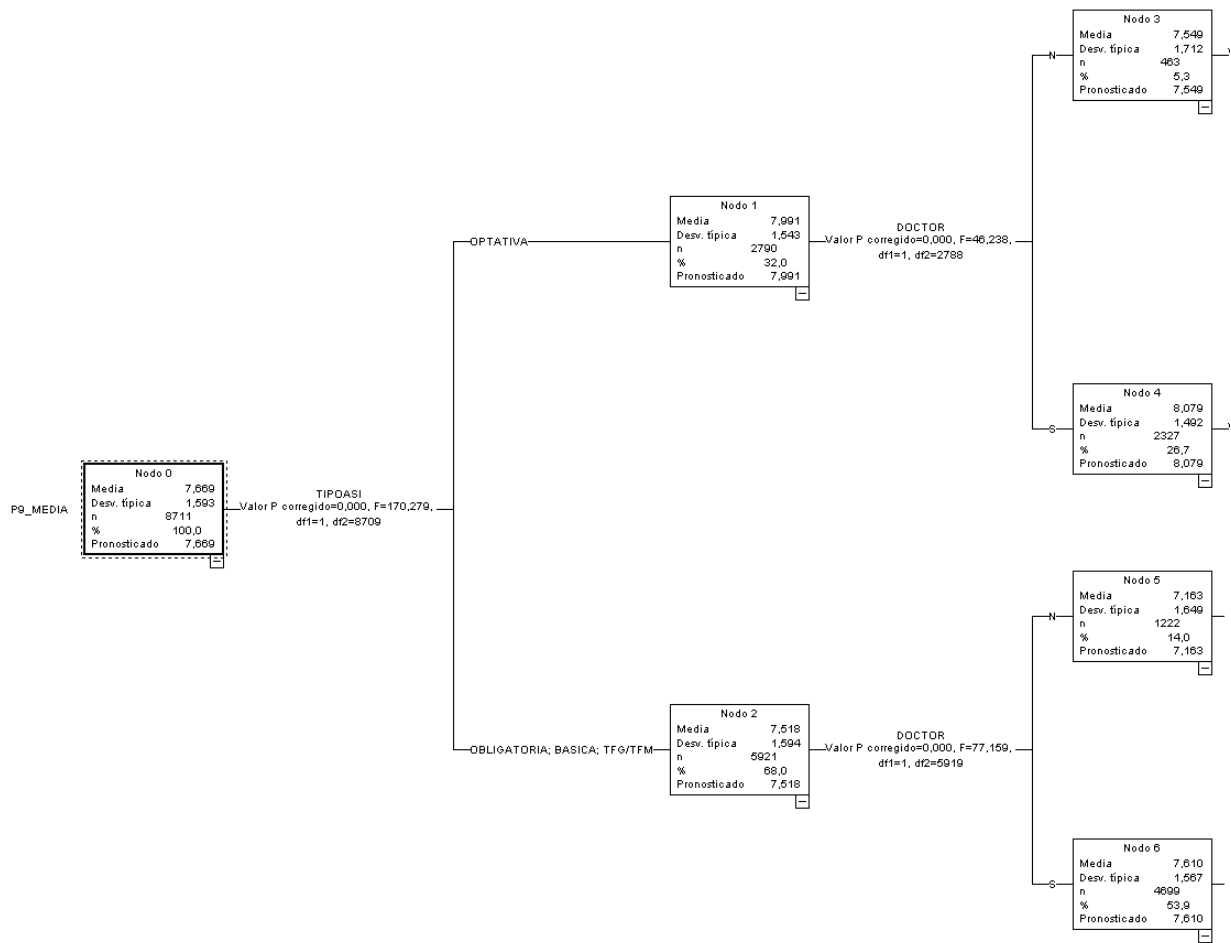


Figura 6: Árbol de decisión siendo la variable dependiente el ítem 9 de la escala. Primera parte

A continuación se presenta la segunda parte del árbol, aquella en que las ramificaciones se han producido con efectos menos significativas de las variables en cuestión:

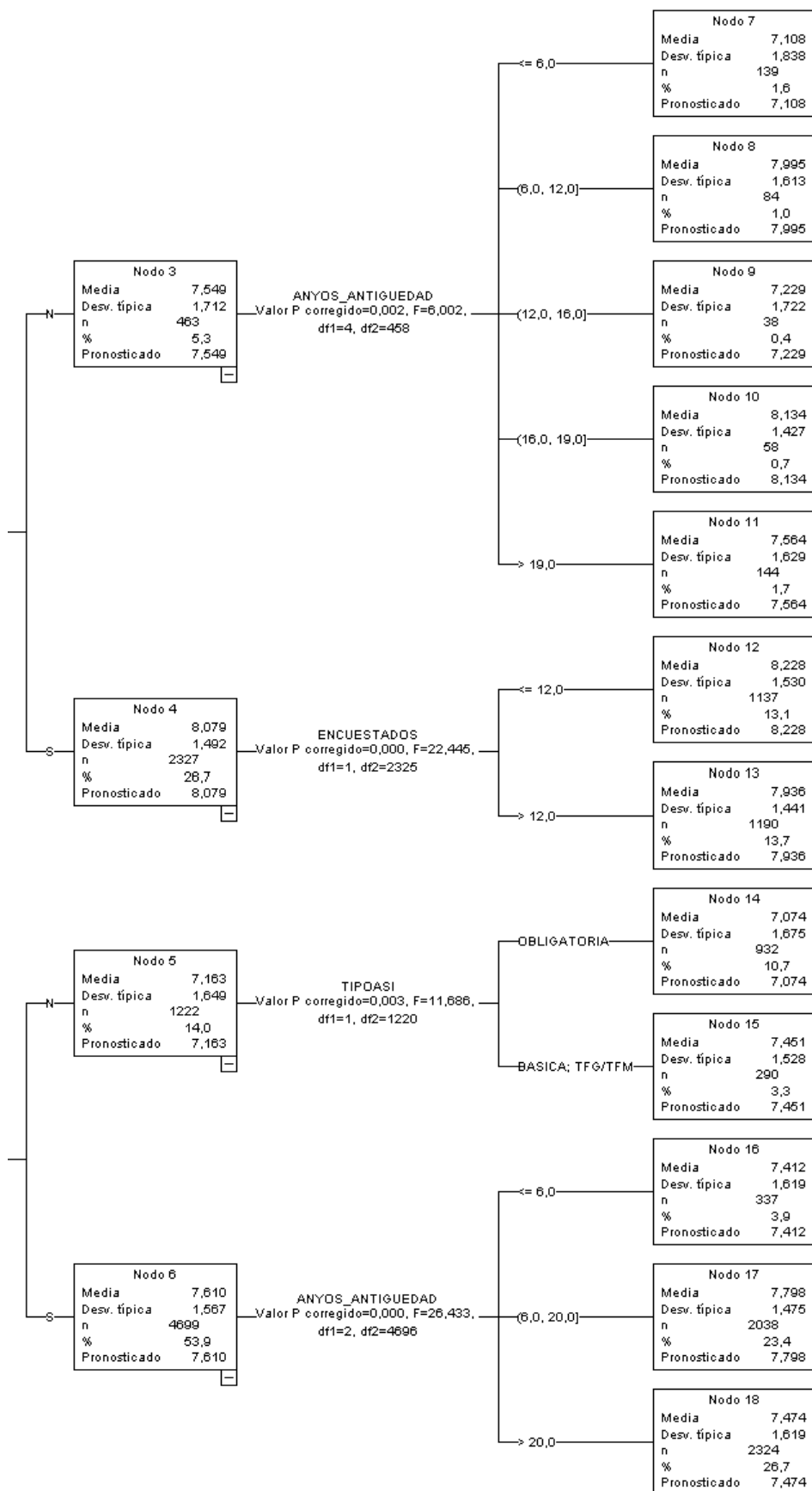


Figura 7: Árbol de decisión siendo la variable dependiente el ítem 9 de la escala. Segunda parte

Se observa que la nota media obtenida es de 7.669, siendo el *Tipo de asignatura* la variable que mayor efecto tiene sobre la satisfacción obtenida en la encuesta, debido a que la primera ramificación se produce en función de esta variable. El valor del estadístico F de esta variable es de 170. En este paso, y con la finalidad de conocer qué variables tienen un efecto más significativo sobre la satisfacción medida por el ítem 9, se calculará el estadístico F para cada una de las variables predictoras cualitativas.

Variable	Estadístico F	p-valor
Tipo asignatura	170.279	0.000
Categoría	71.423	0.000
Tiempo	71.588	0.000
Doctor	133.731	0.000
Sexo	0.341	0.558
Sexenios	120.09	0.000
Tipo titulación	103.423	0.000
Curso	113.994	0.000

**Tabla 20:** Valor del estadístico F sobre el efecto que cada variable tiene sobre el ítem 9 antes de realizar la primera partición

Obtenemos que la variable *Sexo* es la que menos efecto tiene sobre la satisfacción del estudiante, por lo que podemos afirmar que no influye si el profesor es varón o mujer. Por otra parte, se observa que existen varias variables cuyo efecto es significativo, como son el curso, el tipo de titulación o los sexenios, aunque posteriormente no son las más significativas en ninguna de las ramificaciones del árbol. Cabe destacar que el valor del estadístico F ha sido calculado en base al nivel de la variable con más diferencia significativa con el resto frente a estos agrupados en un solo nivel.

Esta división nos permite asegurar que existe una diferencia significativa en la evaluación de la satisfacción en función de si la asignatura es optativa o de otro tipo (Básica, Obligatoria o TFM), siendo las primeras evaluadas, en media, con prácticamente medio punto más. Cabe recordar que las asignaturas optativas son evaluadas por alumnos matriculados en últimos cursos, pues su impartición se concentra en el cuarto curso de grado y el primero de máster.

Por otra parte, la variable *Doctor* será la que provocará la partición tanto en el nodo 1 como en el 2, siendo más significativa en el nodo 2. Es decir, el efecto de ser doctor sobre la satisfacción es más significativo cuando el tipo de asignatura es Básica, Obligatoria o TFM ( $F=77.2$ ) que cuando es Optativa ( $F=46.2$ ). Pese a ello, cabe destacar que la diferencia entre la satisfacción percibida es de medio punto, en media, cuando el profesor que imparte la asignatura es doctor.



La tercera ramificación del árbol se produce para los doctores que cursan las asignaturas del tipo básicas, obligatorias y TFM, pues se percibe una diferencia significativa en función de la variable *Años de antigüedad* ( $F=26.4$ ). En este caso, se observa una diferencia, en media, de 3 décimas entre el grupo de profesores que tiene entre 6 y 20 años de antigüedad en la universidad respecto a los que han trabajado menos de 6 y más de 20 años en la universidad.

A continuación, las asignaturas optativas que son impartidas por un doctor se ramifican, siendo la variable *Encuestados* la que provoca la partición. Se observa como el número de alumnos es significativo en este caso ( $F=22.4$ ), obteniendo, en media, tres décimas menos en la percepción de la satisfacción cuando el grupo tiene más de 12 alumnos.

Las dos últimas particiones se realizan en los nodos de los profesores que no poseen el título de doctor. En primer lugar, tendríamos que existe un efecto significativo del tipo de asignatura cuando hablamos de evaluaciones de la satisfacción en grupos en los que imparte materia un profesor sin doctorado y la asignatura es básica, obligatoria y TFM. Observamos como la percepción de la satisfacción, en media, es prácticamente de 4 décimas superior cuando la asignatura es del tipo básica o TFM. Cabe recordar que las asignaturas básicas se concentran mayormente en el primer curso del grado, representando un 83.86 % de las asignaturas básicas que se imparten en el conjunto de los cursos. Por ello, podría observarse una tendencia de los alumnos de primero de grado a percibir una satisfacción mayor con los profesores no doctores que los alumnos de cursos superiores. Finalmente, tenemos una partición para los profesores que imparten docencia en una asignatura optativa y no son doctores en función de la variable *Años de antigüedad*, que en este caso tiene una  $F=6$ .

Por último, y en base a que disponemos de únicamente 11 variables predictoras y existen efectos de confusión entre dichas variables, destacar que la interpretación de dichos resultados debe realizarse con cautela.

## 5. Conclusiones

El objetivo principal del trabajo ha sido analizar la satisfacción del estudiante con la docencia, así como sentar las bases para posteriores estudios con la finalidad de implementar medidas correctoras a las evaluaciones obtenidas por cada profesor.

Por otra parte, en este trabajo se definieron varios objetivos específicos, siendo el primero de ellos obtener la estructura subyacente de los datos, con las variables de la escala de medida de satisfacción del estudiante como objeto de interés. En este caso se obtuvo que las variables que forman la escala validada por el Instituto de Ciencias de la Educación (2007) pueden agruparse en 3 factores.

Por otro lado, el segundo objetivo planteaba validar la escala de satisfacción del estudiante (Universitat Politècnica de València, 2007). Este trabajo ha demostrado que la escala propuesta por el Instituto de Ciencias de la Educación es la más adecuada y la que presenta mejores índices de bondad de ajuste de entre todos los modelos considerados. En la comparación de modelos realizada, se obtuvo que existen tres modelos entre los que no existen diferencias significativas. Por ello, no se ha propuesto ningún modelo cuyos índices de bondad de ajuste fuesen significativamente mejores que los obtenidos por el modelo propuesto por el Instituto de Ciencias de la Educación. Además, resulta de especial interés la posibilidad de equiparar el segundo factor subyacente (*Satisfacción general*) con una variable observable (*Teniendo en cuenta las limitaciones, pienso que el profesor que imparte esta asignatura debe considerarse un buen profesor*). Por este motivo, se podría utilizar el ítem 9 como medida de satisfacción del alumnado con la docencia.

El tercer objetivo de carácter específico era examinar el efecto sobre la satisfacción del estudiante que podían tener características tanto del profesor como de la asignatura. Para satisfacerlo, se ha utilizado un árbol de decisión. Con esta técnica, se obtuvo que existen diferencias significativas en la satisfacción mostrada por los alumnos en función de si la asignatura es de carácter optativo o de otro tipo (básica, obligatoria o Trabajo Final de Máster). Este hecho confirma la importancia del tipo de asignatura cuando el alumnado expresa la satisfacción, pues es la variable con el efecto más significativo sobre ella. Además, la posesión del título de doctorado también resultaba significativa, siendo los doctores más valorados por los alumnos que los que no ostentaban tal titulación. Aunque su significación era menor, también tenían efecto sobre la satisfacción, en algunos casos, los años de antigüedad, el número de alumnos en la materia y una distinción entre las materias obligatorias y básicas y Trabajo Final de Máster. En base a estos resultados, expertos en la materia podrían utilizar este trabajo como base para proponer diferentes medidas para mejorar la medición de la satisfacción del estudiante.

Este trabajo tiene importantes implicaciones prácticas para los profesores universitarios y órganos de gobierno de las universidades. Esto se debe a que se demuestra que la utilización de la encuesta actual por parte de la universidad es válida y fiable, pues mide aquello que se pretende.

Asimismo, aporta un nuevo enfoque de carácter cuantitativo a investigaciones precedentes sobre el uso de escalas de medición en educación universitaria.

A continuación, se hará una valoración sobre las limitaciones que hemos tenido en este trabajo. Sin duda, uno de los mayores problemas con las que cuenta es no poder plantear un modelo multinivel, pues, según lo establecido por Muthén (1997), sería necesario disponer de 15 evaluaciones por profesor, condición que está lejos de satisfacerse. Por otra parte, nuestra base de datos cuenta solo con evaluaciones del curso 2016-2017, por lo que la validación se realizó con un curso académico. Por último, el árbol de decisión se obtuvo en base a 11 variables predictoras, no estando disponibles otras características del proceso de enseñanza-aprendizaje que podrían tener un efecto significativo sobre la satisfacción.

Finalmente, realizaremos diferentes propuestas que nos permitirían profundizar en futuras líneas de investigación. En primer lugar, podríamos considerar la recopilación de datos de distintos cursos académicos con la finalidad de obtener suficiente información sobre los profesores. Con esto, seríamos capaces de proponer un modelo multinivel. Además, debido al perfil de esta universidad, solo disponemos de evaluaciones realizadas por alumnos de la rama científica, bellas artes y Administración y Dirección de Empresas (ADE). Por ello, resultaría interesante comparar los resultados obtenidos con los que proporcionan estudiantes de otras ramas (Salud, Humanidades,...), pues podrían tener un distinto grado de madurez o preocupaciones diametralmente opuestas que les llevaran a evaluar de forma distinta la satisfacción con la docencia. Además, para obtener un árbol de decisión más preciso, se considera que necesitaríamos más variables sobre las características del profesor y la asignatura. Cabe destacar que se podrían añadir ítems a la encuesta con la finalidad de obtener mediciones sobre cuestiones que se consideran relevantes para estudiar los posibles factores que inciden sobre la satisfacción. Esta manera de proceder es la que llevó a la inclusión de un décimo ítem, el cual no forma parte de la escala propuesta por el Instituto de Ciencias de la Educación, y que no se utiliza para medir la satisfacción del estudiante, sino para medir la actividad docente.

Por todos estos motivos, este trabajo es un punto de partida para, mediante futuras líneas de investigación, mejorar la satisfacción del estudiante y conocer qué variables asociadas al profesorado tienen un gran efecto sobre dicha satisfacción de manera más precisa.

## Referencias

- [1] 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.
- [2] Arruda, J. E., Weiler, M. D., Valentino, D., Willis, W. G., Rossi, J. S., Stern, R. A., Gold, S.M. & Costa, L. (1996). A guide for applying principal-components analysis and confirmatory factor analysis to quantitative electroencephalogram data. *International Journal of Psychophysiology*, 23(1-2), 63-81.
- [3] Bentler, P.M. (2005). *EQS 6 structural equations program manual*. Encino, California: Multivariate Software (www.mvsoft.com)
- [4] Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological bulletin*, 88(3), 588-606.
- [5] Biggs, D., Ville, B. y Suen, E. (1991). A method of choosing multiway partions for classification and decision tress. *Journal of Applied Statistics*, 18(1), 49 - 60.
- [6] Bollen, K. A. (1989). A new incremental fit index for general structural equation models. *Sociological Methods & Research*, 17(3), 303-316.
- [7] Browne, M.W. y Cudeck, R. (1993). Alternative ways of assessing model fit. In K.A. Bollen y J.S. Long(Eds.), *Testing Structural Equation Models*, (pp. 136-162).
- [8] Byrne, B. M. (2013). *Structural equation modeling with EQS: Basic concepts, applications, and programming*. Routledge.
- [9] Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological bulletin*, 56(2), 81.
- [10] Carretero-Dios, H., & Pérez, C. (2005). Normas para el desarrollo y revisión de estudios instrumentales. *International Journal of clinical and health psychology*, 5(3),521-551.
- [11] Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural equation modeling*, 9(2), 233-255.
- [12] Comrey, A. L. (1988). Factor-analytic methods of scale development in personality and clinical psychology. *Journal of consulting and clinical psychology*, 56(5), 754-761.
- [13] Conchado, A. (2011). *Modelización multivariante de los Procesos de Enseñanza-Aprendizaje basados en Competencias en Educación Superior*. Published doctoral dissertation, Universitat Politècnica de València

- [14] Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of applied psychology*, 78(1), 98-104.
- [15] Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334.
- [16] Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological bulletin*, 52(4), 281-302.
- [17] Eberts, R., Hollenbeck, K., & Stone, J. (2002). Teacher performance incentives and student outcomes. *Journal of human resources*, 913-927.
- [18] Escudero, T. (2006). La evaluación del profesorado en España en la última década: Nuevos retos. *Encuentro*, 10, 1996-2006. Organizadas por ADECA y Universidad de Zaragoza.
- [19] Escudero, T., Pino, J.L. y Rodríguez, C. et. al. (2006). *Metaevaluación externa del Plan de Complementos Retributivos de carácter individualizado, del Personal Docente e Investigador de la Universidad de La Rioja*.
- [20] Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological assessment*, 7(3), 286-299.
- [21] Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of marketing research*, 39-50.
- [22] García Valcárcel, A. (1991). El comportamiento de los profesores universitarios en el aula. *Studia Pedagógica*, 23, 135-153.
- [23] García-Valcárcel, A. (1992). Características del buen profesor universitario según estudiantes y profesores. *Revista de Investigación Educativa*, 19, 31-50
- [24] Henry, G.T. y Rubenstein, R.(2002). Paying for grades: Impact of merit-bases financial aid on educational quality. *Journal of Policy Analysis and Management*,21(1), 93-109.
- [25] Hox, J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, New Jersey: Lawrence Erlbaum & associates.
- [26] Hoyle, R. H. (1995). *Structural equation modeling: Concepts, issues, and applications*. Sage Publications.
- [27] Hu, L. T., & Bentler, P. M., (1995). Evaluating model fit. *Structural equation modeling: Concepts, issues, and applications*, 76-99.
- [28] Instituto de las Ciencias de la Educación. (1994). *Informe sobre la evaluación del profesorado*. ICE -Universidad Politécnica de Valencia.

- [29] James, L. R., Mulaik, S. A., & Brett, J. M. (1982). *Causal analysis*. Beverly Hills: Sage Publications
- [30] Jöreskog, K.G. & Sörbom, D. (1989). *LISREL 7: User's reference guide*. Chicago, Illinois: Scientific Software International.
- [31] Jornet, J., Villanueva, P., Suárez, J. M., & Alfaro, I. (1988). Proyecto de implantación de un sistema de evaluación del profesorado en la Universidad de Valencia. Universidad de Valencia. *Consideraciones metodológicas sobre la evaluación y mejora de la docencia universitaria*. Informes de Investigación Evaluativa, 1.
- [32] Kaplan, D. (1998). Methods for multilevel data analysis. In G.A. Marcoulides (Eds.), *Modern methods for business research*(pp.337-358). Mahwah, New Jersey: Lawrence Erlbaum & associates.
- [33] Kass, G.V. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics*, 29(2), 119-127.
- [34] Kenny, D. A. (1979). *Correlation and causality*. New York: Wiley
- [35] Lévy-Mangin, J.P., & Varela, J. (2006). *Modelización con estructuras de covarianzas en ciencias sociales. Temas esenciales, avanzados y aportaciones especiales*. A Coruña: Netbiblo.
- [36] Likert R.A. (1932). A technique for development of attitude scales. *Archives of Psychology*, 140, 44-53.
- [37] Llanos, F., Rosas, A., Mendoza, D., & Contreras, C. (2001). Comparación de las escalas de Likert y Vigesimal para la evaluación de satisfacción de atención en un hospital del Perú. *Revista Medica Herediana*, 12(2), 52-57.
- [38] Long, J.S. (1983). *Confirmatory Factor Analysis*. Newbury Park, California: Sage Publications
- [39] Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57(3), 519-530.
- [40] Marsh, H. W., & Overall, J. U. (1977). The Relationship Between Students' Evaluations of Faculty and Instructional Improvement.
- [41] Marsh, H. W., & Overall, J. U. (1980). Validity of students' evaluations of teaching effectiveness: Cognitive and affective criteria. *Journal of Educational Psychology*, 72(4), 468-475.

- [42] Martínez, M. (2008). *Estudio del cuestionario de evaluación del profesorado de la UPV mediante opinión de los estudiantes. Tratamiento estadístico*. Published doctoral dissertation, Universitat Politècnica de València.
- [43] Mas, Ò. (2011). El profesor universitario: sus competencias y formación. Profesorado. *Revista de Currículum y Formación de Profesorado*, 15(3), 195-211.
- [44] McDonald, R. P., & Marsh, H. W. (1990). Choosing a multivariate model: Noncentrality and goodness of fit. *Psychological bulletin*, 107(2), 247-255.
- [45] Mehrens, W. A. (1990). Combining evaluation data from multiple sources. In J. Millman y L. Darling-Hammond (Eds.), *The New Handbook of Teacher Evaluation. Assessing Elementary and Secondary School Teachers* (pp. 322-334). Newbury Park, California: Sage Publications
- [46] Milanowski, A. (2007). Performance pay systems preferences of students preparing to be teachers. *Education Finance and Policy*, 2(2), 111-132.
- [47] Mizala, A. y Romaguera, P. (2004). School and teacher performance incentives: The Latin American experience. *International Journal of Educational Development*, 24(6), 739-754.
- [48] Muthén, B.O. (1994). Multilevel covariance structure analysis. *Sociological Methods & Research*, 22, 376-398.
- [49] Muthén, B.O. (1997). Latent variable modeling of longitudinal and multilevel data. In A.E. Raftery (Eds.), *Sociological Methodology 1997* (pp.453-481). Washington, DC : American Sociological Association
- [50] Nunnally, J. C., & Bernstein, I. J. (1995). *Teoría psicométrica*. Madrid:McGraw-Hill.
- [51] Pérez, C. (2011). *Técnicas de análisis multivariante de datos. Aplicaciones con SPSS*. Madrid: Pearson educación.
- [52] Pérez-Gil, J. A., Chacón, S., & Moreno, R. (2000). Validez de constructo: el uso de análisis factorial exploratorio-confirmatorio para obtener evidencias de validez. *Psicothema*, 12(2), 442-446.
- [53] Satorra, A. y Bentler, P.M. (1988). *Scaling corrections for statistics in covariance structure analysis*. Los Angeles: UCLA statistics series 2.
- [54] Tejedor, F. J. (1998). *Los alumnos de la Universidad de Salamanca. Características y rendimiento académico* (Vol. 34). Universidad de Salamanca.
- [55] Tejedor, F. J. (2003). Un modelo de evaluación del profesorado universitario. *Revista de investigación educativa*, 21 (1), 153-182.

- 
- [56] Tejedor, F. J., Castro, C., & García, C. (1988). Evaluación del profesorado universitario por los alumnos de la Universidad de Santiago (Santiago de Compostela, Informe).
- [57] Tejedor, F. J., & García-Valcárcel, A. (1996). La evaluación de la calidad de la docencia universitaria (en el marco de la evaluación institucional) desde la perspectiva del alumno. *Evaluación educativa II. Evaluación institucional*, 93-122.
- [58] Tejedor, F. J., & Jornet Meliá, J. M. (2008). La evaluación del profesorado universitario en España. *Revista electrónica de investigación educativa*, 10, 1-29.
- [59] Tejedor, F. J., & Montero, M. L. (1990). Indicadores de la calidad docente para la evaluación del profesor universitario. *Revista española de pedagogía*, 48(186),259-279.
- [60] Trent, J. W., & Cohen, A. M. (1973). Research on teaching in higher education. In R.M. Travers (Eds.), *Second handbook of research on teaching* (pp.997-1071). Chicago.
- [61] Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38(1), 1-10.
- [62] Universitat Politècnica de València. (2007). *Manual de evaluación de la actividad docente del profesorado de la Universidad Politécnica de Valencia*.



## 6. Anexo

En el apartado 4.4 (*Análisis exploratorio sobre los posibles factores que inciden en la satisfacción*), calculamos el estadístico F de cada variable cualitativa para conocer la significación que tenía sobre la variable dependiente, el ítem criterio de la escala (*Buen profesor*). Para ello, se realizó un ANOVA para cada variable predictora cualitativa como factor y el ítem 9 como variable dependiente. Tras hallar qué nivel de cada factor tenía más diferencias significativas con el resto, se agrupaban el resto de niveles frente a él, con la finalidad de realizar tal partición del árbol. Tras realizar otro ANOVA, se halló el valor del estadístico F para cada variable cualitativa. En estos análisis, obtuvimos los siguientes intervalos LSD:

1. En los intervalos LSD de la variable *Doctor* S representa a los que poseen el título de doctorado y N a los que no. Se observa como existe una diferencia significativa en media, pues los intervalos no se solapan, la cual alcanza, en media, las 5 décimas.

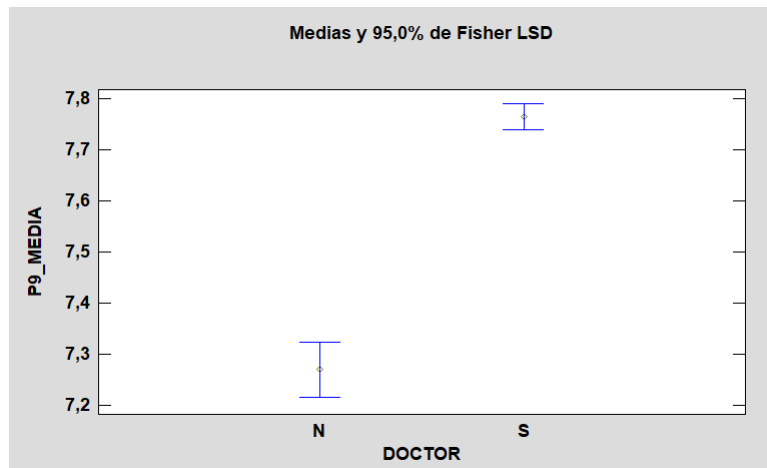


Figura 8: Intervalos LSD de la variable *Doctor*

2. Por otro lado, tenemos los intervalos LSD de la variable *Curso*, en los cuáles se observó que las diferencias más significativas se encuentran entre los primeros cuatro años de grado (representados por el uno) y los años de máster, juntamente con el quinto curso de Arquitectura (nivel 0). En este caso, en el segundo grupo de cursos se obtiene una satisfacción del estudiante, en media, 4 décimas superior a la obtenida en los primeros cuatro años de grado.

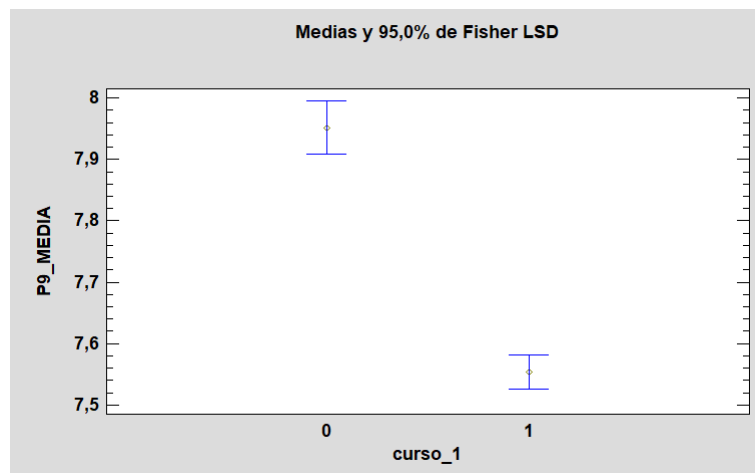


Figura 9: Intervalos LSD de la variable *Curso*

3. También es una variable de interés el *Tipo de titulación*, el cual diferencia entre los títulos de Grado (G) y Posgrado (P). En este caso, la diferencia mostrada, en media, en la satisfacción con la docencia prácticamente es de 4 décimas, valorando más positivamente los alumnos de máster.

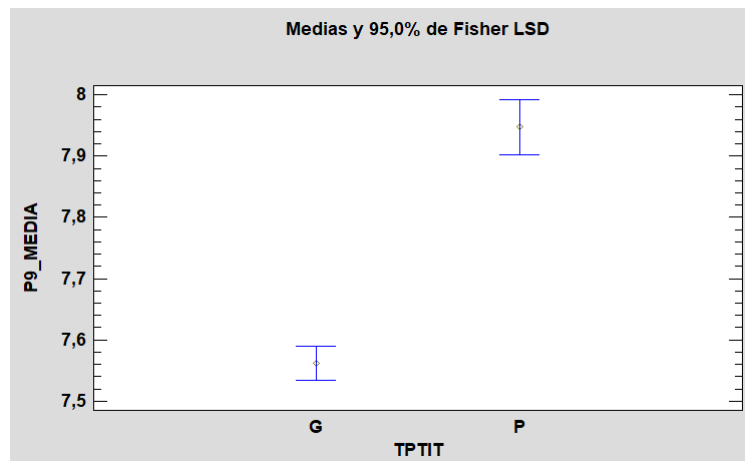


Figura 10: Intervalos LSD de la variable *Tipo de titulación*

4. En la variable *Sexenios* la diferencia vuelve a alcanzar, en media, prácticamente las 4 décimas, siendo estas favorables al nivel 0 (aquellos que tienen, al menos, un sexenio reconocido) frente al nivel 1 (profesores que no cuentan con ningún sexenio).

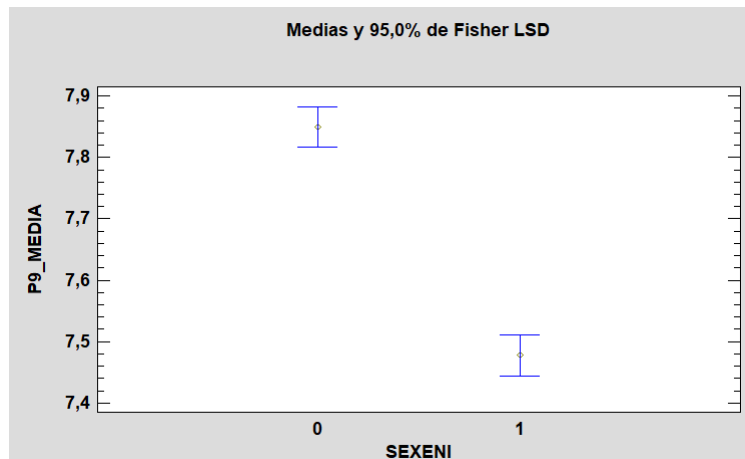


Figura 11: Intervalos LSD de la variable *Sexenios*

5. La variable *Sexo* es la única que no presenta ningún efecto significativo, por menor que sea, sobre la satisfacción general. Esto es observable al comprobar que los intervalos LSD para ambos niveles se solapan.

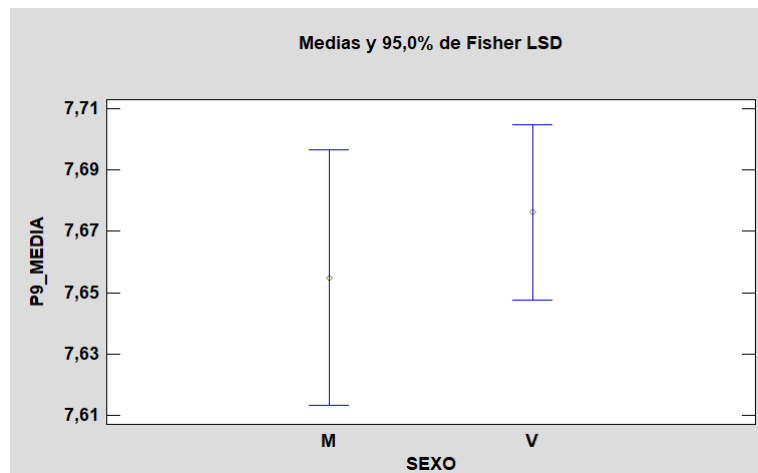


Figura 12: Intervalos LSD de la variable *Sexo*

6. El *Tiempo de dedicación* también resulta significativo, siendo 35 centésimas la diferencia, en media, entre los profesores que trabajan a tiempo completo en la universidad (TC) frente los que lo hacen a tiempo parcial (TP).

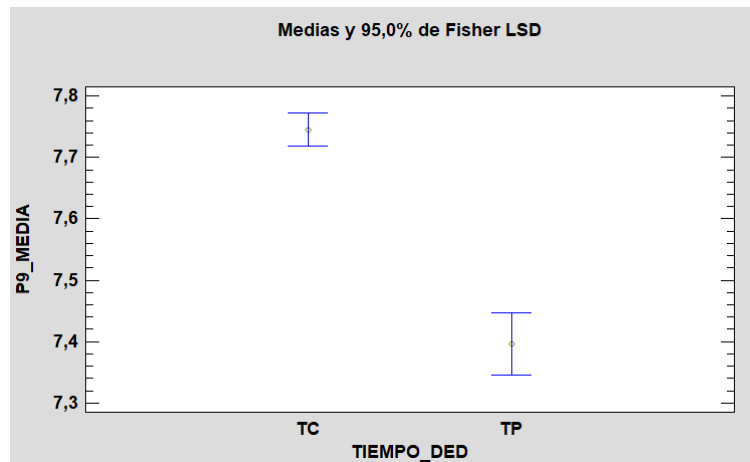


Figura 13: Intervalos LSD de la variable *Tiempo de dedicación*

7. Por otro lado, se encuentra la variable *Categoría docente*, la cual presenta la diferencia más significativa entre sus niveles entre los asociados (nivel 1) y el resto de categorías presentes en la universidad (nivel 0). En este caso, existe una diferencia, en media, de la satisfacción percibida de los alumnos de 35 centésimas en favor del segundo grupo.

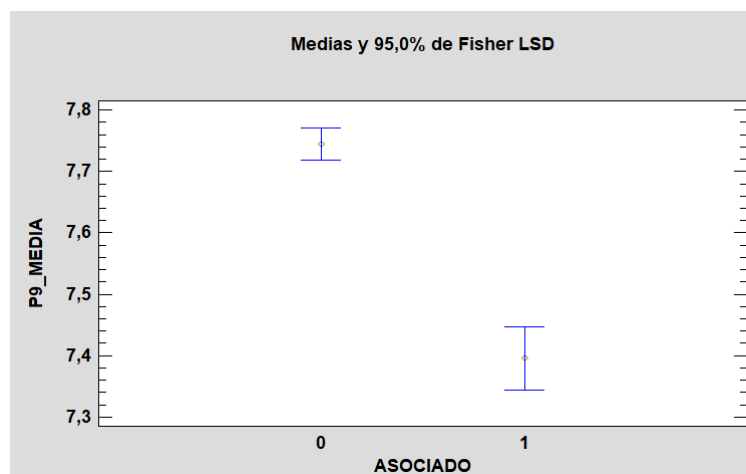


Figura 14: Intervalos LSD de la variable *Categoría docente*

8. Finalmente, tenemos el *Tipo de asignatura*. El nivel 0 engloba las asignaturas que son de carácter *Básico, Obligatorio o TFM*, mientras que el nivel 1 representan las *Optativas*. Este resultado es la variable cuyo efecto es más significativo (valor del estadístico F mayor), y la diferencia en media es de 5 décimas en favor de los profesores que imparten una materia optativa frente al resto.

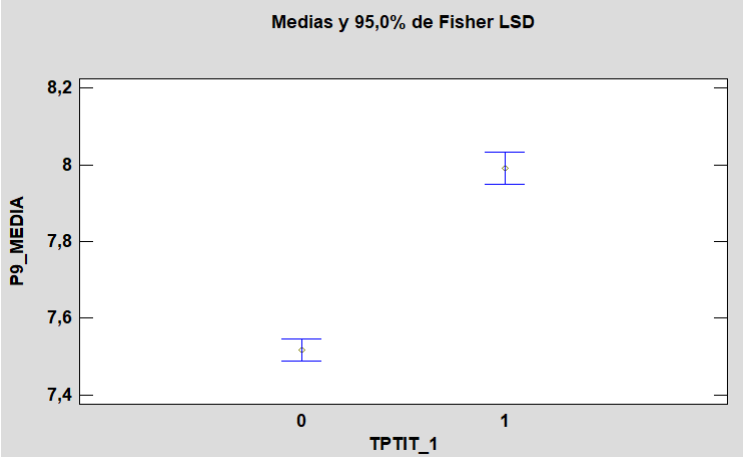


Figura 15: Intervalos LSD de la variable Tipo de asignatura