

Taborga, Valdivieso; Eduardo, Carlos

## Article

# Comparación de los modelos formativo, reflexivo y de antecedentes de evaluación estudiantil del servicio de docencia

Revista de Métodos Cuantitativos para la Economía y la Empresa

## Provided in Cooperation with:

Universidad Pablo de Olavide, Sevilla

*Suggested Citation:* Taborga, Valdivieso; Eduardo, Carlos (2013) : Comparación de los modelos formativo, reflexivo y de antecedentes de evaluación estudiantil del servicio de docencia, Revista de Métodos Cuantitativos para la Economía y la Empresa, ISSN 1886-516X, Universidad Pablo de Olavide, Sevilla, Vol. 16, pp. 95-120

This Version is available at:

<https://hdl.handle.net/10419/113863>

### Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

### Terms of use:

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<http://creativecommons.org/licenses/by-sa/3.0/es/>



## Comparación de los modelos formativo, reflexivo y de antecedentes de evaluación estudiantil del servicio de docencia

VALDIVIESO TABORGA, CARLOS EDUARDO

Departamento de Ciencias Exactas

Universidad Privada Boliviana, UPB (Bolivia)

Correo electrónico: [cvaldivieso@upb.edu](mailto:cvaldivieso@upb.edu)

### RESUMEN

Se ha llevado a cabo un análisis comparativo entre un modelo con enfoque de medición formativo, otro reflexivo y un tercero de antecedentes, de las dimensiones de la calidad de servicio de docencia.

Previamente se diseñaron instrumentos válidos y fiables para la medición de la calidad de servicio y sus constructos relacionados, a través de un estudio exploratorio para obtener validez de contenido. Además, se llevaron a cabo un análisis factorial exploratorio (AFE) y otro confirmatorio (AFC), para obtener la estructura dimensional adecuada e ítems relevantes.

Usando modelos de estructuras de covarianza (MEC) se comprobó que el modelo de antecedentes, aunque presenta igual poder predictivo, tiene mejores índices de ajuste que los modelos reflexivo y formativo.

**Palabras clave:** calidad de servicio de docencia; valor percibido; satisfacción estudiantil; intenciones de comportamiento; modelos estructurales; enfoque formativo y reflexivo.

**Clasificación JEL:** M390; I230.

**MSC2010:** 62H25; 90B60.

# Comparison of Formative, Reflective, and Antecedents Models of Students Evaluation of Teaching Service

## ABSTRACT

A comparative analysis between three measurement models (with formative, reflective, and antecedents approach) has been conducted to measure the dimensions of the teaching service quality.

Previously valid and reliable instruments have been designed to measure the service quality and related constructs. An exploratory study has been performed to obtain content validity. Additionally, both exploratory and confirmatory factor analysis (EFA and CFA, respectively) have been conducted to obtain the appropriate dimensional structure and relevant items.

Using structural equations modeling (MEC) has been found that the antecedent model, although it has the same predictive power, has better fit indices than reflective and formative models.

**Keywords:** teaching service quality; perceived value; student satisfaction; behavioral intentions; structural models; formative and reflective approach.

**JEL classification:** M390; I230.

**MSC2010:** 62H25; 90B60.



## 1. INTRODUCCIÓN

En Economía y Empresa, es vital el uso de indicadores multidimensionales para evaluar fenómenos más o menos subjetivos; donde no existe un consenso es en qué tipo de indicadores utilizar. Domínguez *et al.* (2011) proporcionan un resumen de las diferentes formas de calcular indicadores sintéticos, con sus principales ventajas e inconvenientes; Fedriani y Martín (2011) proponen un conjunto de indicadores multidimensionales no necesariamente sintéticos y aplicables a las Ciencias Sociales, ejemplificados en la evaluación de la pobreza; Poza y Fernández (2011) también miden la pobreza desde la vertiente multidimensional, pero utilizando ecuaciones estructurales.

Algunos investigadores (e.g. Diamantopoulos, 2008 y Wilcox *et al.*, 2008) sostienen que la generalidad de los estudios en el campo del *management* se ha fundamentado en planteamientos que asumen la dirección de causalidad desde el constructo hacia sus medidas (indicadores reflexivos); señalando que este error puede dar lugar a problemas de una especificación inadecuada de los modelos y tener consecuencias adversas para la validez de sus conclusiones, además de originar diferencias en los valores de las rutas establecidas en modelos estructurales (Jarvis *et al.*, 2003; Podsakoff *et al.*, 2006). Las investigaciones acerca de la medición de la calidad de servicio y sus dimensiones, han adoptado una especificación de medición reflexiva (e.g. Dabholkar *et al.*, 1996 y Alén y Fraiz, 2006). Para probar si es mejor un enfoque formativo, se necesita incluir a la calidad de servicio en un modelo estructural de relaciones entre otros constructos (Bollen, 1989).

Varias investigaciones han concluido que las actitudes de la calidad y los juicios de la satisfacción juegan un papel en la formación de las intenciones de compra futuras (Cronin y Taylor, 1992; Taylor y Baker, 1994). Otros trabajos han incluido en esta relación al valor percibido como antecedente de las intenciones de compra (Brady *et al.*, 2005; Choi *et al.*, 2004). Sin embargo, sigue siendo desconocido si estos resultados pueden ser generalizados a los consumidores de países de América Latina. Si bien el modelo de la satisfacción, en el cual la satisfacción actúa como mediador entre la calidad de servicio y las intenciones de comportamiento, ha sido estudiado por varios investigadores (Dabholkar *et al.*, 2000; Brady *et al.*, 2005), éste solo ha sido aplicado a la Educación Superior por Banwet y Datta (2003). Tampoco existen investigaciones comparativas entre modelos con enfoque de medición formativo y reflexivo en este contexto. Por otro lado, algunos investigadores (Dabholkar *et al.*, 2000 y Wang y Lo, 2002) plantearon un modelo en el que las dimensiones de la calidad de servicio son consideradas antecedentes del constructo más que componentes, en el cual la satisfacción es un mediador poderoso entre la calidad de servicio y las intenciones de comportamiento, concluyendo que su modelo tenía mejores índices de ajuste que el de componentes.

El propósito de esta investigación es realizar un análisis comparativo de tres modelos (el reflexivo, el formativo y el de antecedentes) a fin de probar cuál tiene mejor especificación para medir las dimensiones de la calidad de servicio de docencia, dentro de un modelo estructural de relación con el valor percibido, la satisfacción y las intenciones de comportamiento de postdocencia, y comprobar el rol mediador de la satisfacción.

## **2. REVISIÓN DE LITERATURA**

### **2.1. La investigación sobre la medición de la calidad de servicio**

Grönroos (1984) remarcó la necesidad de una investigación de la calidad de servicio desde el punto de vista de los clientes, planteando un modelo de dos dimensiones: la técnica (resultado) y la funcional (proceso). Hasta el año 1985, pocos investigadores intentaron modelar la calidad de servicio, debido a la dificultad de delimitarla y medirla. La medición cuantitativa tiene su inicio en los estudios que realizaron Parasuraman *et al.* (1985; 1988), dando como resultado un instrumento genérico denominado SERVQUAL, el cual sería refinado posteriormente por los mismos investigadores (Parasuraman *et al.*, 1991). Este instrumento “operacionaliza” la calidad de servicio en 5 dimensiones (tangibles, confiabilidad, seguridad, empatía y receptividad) y mide la deficiencia entre las expectativas de los clientes con sus percepciones del servicio.

La medición de la calidad de servicio por medio del modelo de brecha y el SERVQUAL (Parasuraman *et al.*, 1985; 1988), fue criticado tanto por Cronin y Taylor (1992) como por Teas (1993), que propusieron el SERVPERF (que mide solamente las percepciones) y el modelo de desempeño evaluado (que pondera los atributos de la calidad de servicio según su nivel de expectativa), respectivamente. Posteriormente varios investigadores han llevado a cabo trabajos empíricos de replicación, adaptación o refinación del SERVQUAL en distintos servicios (Carman, 1990; 2000; Boulding *et al.*, 1993). Cronin y Taylor (1992) afirmaron que la calidad de servicio es un antecedente de la satisfacción, la cual influye significativamente en las intenciones de compra, surgiendo el modelo de calidad de servicio percibido y satisfacción (Spreng y Mackoy, 1996). Luego Dabholkar *et al.* (2000), examinaron la relación entre la calidad de servicio, la satisfacción y las intenciones de comportamiento, proponiendo el modelo de antecedentes y mediador. Cronin y Taylor (1992) destacaron la importancia del “valor percibido” por el cliente y, a partir de entonces, los investigadores han propuesto el modelo del valor percibido para el mejoramiento y entendimiento de la calidad de servicio (e.g. Oh, 1999 y Choi *et al.*, 2004).

La investigación actual está enfocada en la identificación de los factores que “conducen” las compras de los consumidores de servicios: calidad de servicio, valor del servicio, satisfacción del cliente e intenciones de comportamiento (Parasuraman *et al.*, 1988; Zeithaml *et al.*, 1996). Pocos investigadores han analizado estos constructos usando modelos comprensivos; es decir que examinen todas las relaciones entre ellos (Brady *et al.*, 2005).

### **2.2. La calidad de servicio como un constructo unidimensional**

La estrategia de conceptualización unidimensional de la calidad de servicio es eficaz y sencilla, pero no puede discernir su naturaleza compleja y multifacética. Varios investigadores han medido la calidad de servicio como un constructo unidimensional. La mayoría (Boulding *et al.*, 1993; Cronin y Taylor, 1992; Zeithaml *et al.*, 1996) han usado una medida simple de un ítem que hace imposible el cálculo de la fiabilidad. Otros estudios han usado medidas multi-ítem (Dabholkar *et al.*, 1996; Brady *et al.*, 2002).

### 2.3. La calidad de servicio como un constructo multidimensional

Los modelos multidimensionales de la calidad de servicio ayudan a los investigadores a discernir su naturaleza compleja; sin embargo, pueden ser incompletos, ya que si se conceptualiza la calidad de servicio en un nivel abstracto, el enfoque multidimensional no diferencia la calidad de servicio en sí de sus componentes o dimensiones. Las dimensiones más usadas han sido: tangibles, confiabilidad, seguridad, empatía, receptividad y resultado.

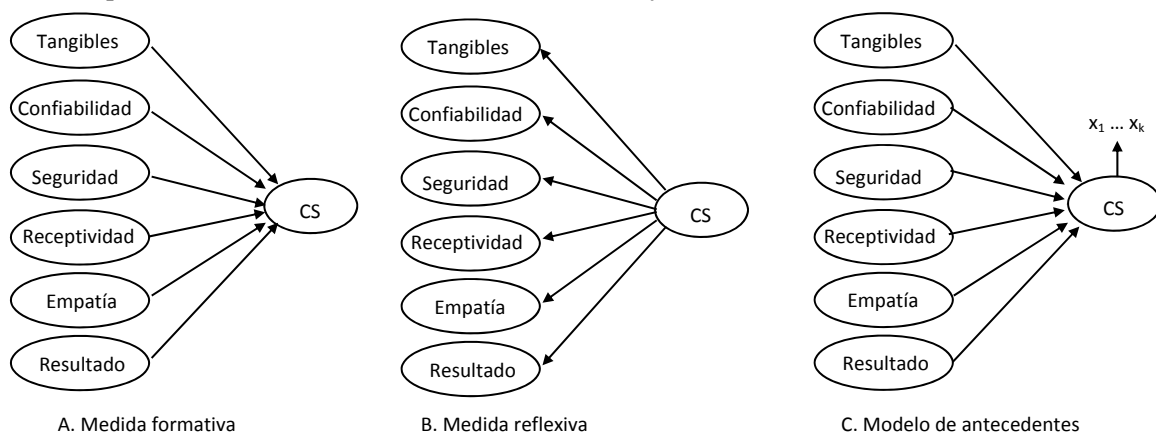
Asubonteng (1996) y Ladhari (2008) manifiestan que la mayoría de las escalas para medir la calidad de servicio son multidimensionales, con variación del número de dimensiones (desde 2 hasta 10), de acuerdo al contexto de servicio.

### 2.4. Modelos de medición formativa, reflexiva y de antecedentes, de la calidad de servicio

#### 2.4.1. Diferencias conceptuales

La característica de un modelo de medición reflexivo es que un cambio en la variable latente (no observable) será reflejado en un cambio en todos sus indicadores. En el modelo formativo, son los indicadores los que causan una variable latente, así que un cambio en ella no es acompañado por un cambio en todos sus indicadores. La diferencia entre los dos enfoques está en la prioridad causal entre la variable latente y sus indicadores (Bollen, 1989). En la Figura 1 se presentan las diferencias entre una especificación de medición reflexiva, formativa y de antecedentes de la calidad de servicio.

Figura 1. Especificación de medición reflexiva, formativa y de antecedentes de la calidad de servicio (CS)

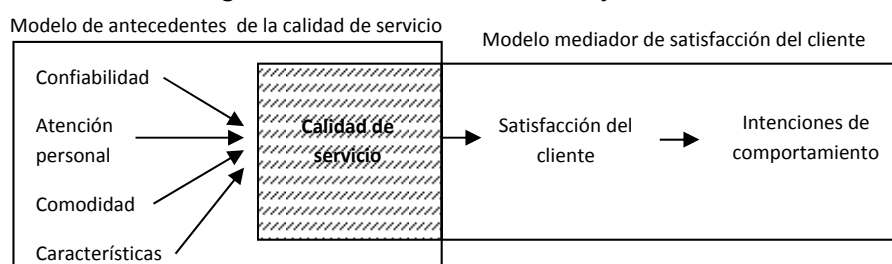


En el modelo formativo, las variables observadas causan la variable latente (panel A, a la izquierda de la Figura 1). Las dimensiones de tangibles, confiabilidad, seguridad, receptividad, empatía y resultado, causan el constructo de la calidad de servicio de docencia. En el modelo reflexivo la variable latente causa las variables observadas (panel B, en el centro de la Figura 1). Las dimensiones son un reflejo de la calidad de servicio de docencia. En ambos modelos la operacionalización de la calidad de servicio es multidimensional. En el modelo de antecedentes (panel C, a la derecha de la Figura 1), las dimensiones son constructos antecedentes de la calidad de servicio, y la operacionalización del constructo se realiza desde una perspectiva unidimensional, que puede ser medida por un ítem o por varios ítems ( $x_1 \dots x_k$ ).

Si bien el uso de medidas multi-ítem ha aumentado indudablemente la calidad metodológica de los estudios empíricos (desde el paradigma de Churchill, 1979), también ha llevado a una adherencia ciega a la perspectiva de medición reflexiva. Hasta ahora, no existen investigaciones que hayan planteado la medición de las dimensiones de la calidad de servicio, con indicadores formativos.

Dabholkar *et al.* (2000) y Wang y Lo (2002) plantearon un modelo en el que las dimensiones de la calidad de servicio son sus antecedentes y en el que la satisfacción es un mediador entre la calidad de servicio y las intenciones de comportamiento (véase la Figura 2). Una variable mediadora es aquella que interactúa con una variable predictora para conjuntamente afectar la variable dependiente de alguna manera (Taylor, 1997).

Figura 2. Modelo de antecedentes y mediador



Fuente: Dabholkar *et al.* (2000)

Dabholkar *et al.* (2000) y Wang y Lo (2002) probaron que el modelo de antecedentes tiene mejores índices de ajuste que el de componentes (reflexivo) y que la satisfacción actúa como un mediador poderoso. La única investigación de un modelo de antecedentes en la Educación Superior es la de Banwet y Datta (2003), con su modelo de la calidad de servicio de docencia.

#### 2.4.2. Criterios para la elección entre modelos reflexivos y formativos

La elección del modelo de medición más apropiado puede ser difícil. Se han formulado cuatro criterios que pueden ayudar en la elección de un modelo formativo. La pregunta clave es si cada dimensión hace una contribución única a la comprensión del constructo, y para constatar este hecho, las condiciones son: 1) deben ser distinguibles unos de otros; 2) no deben ser conceptualmente intercambiables; 3) no deben covariar a un alto nivel empíricamente; y 4) no deben tener redes nomológicas similares, es decir, no deben tener los mismos constructos antecedentes y consecuentes (Jarvis *et al.*, 2003; Coltman *et al.*, 2008).

### 3. MÉTODOS ESTADÍSTICOS PARA EL ANÁLISIS DE DATOS

#### 3.1. Análisis factorial

Es una técnica en la cual los factores son seleccionados para explicar las interrelaciones entre variables. Las variables originales juegan el papel de variables dependientes que se explican por factores comunes y únicos, que no son observables. Puede ser exploratorio (no se conoce *a priori* el número de factores y se determina en la aplicación empírica) o confirmatorio (los factores están fijados *a priori*, utilizándose contrastes empíricos para su corroboración).

### 3.1.1. Análisis Factorial Exploratorio (AFE)

#### 3.1.1.1. Especificación del modelo

Para la formulación del modelo se considerará que las variables observables  $X_1, X_2, \dots, X_p$  son variables tipificadas (Uriel y Aldás, 2005). El modelo del análisis factorial es el siguiente:

$$\begin{cases} X_1 = l_{11}F_1 + l_{12}F_2 + \dots + l_{1m}F_m + e_1 \\ X_2 = l_{21}F_1 + l_{22}F_2 + \dots + l_{2m}F_m + e_2 \\ \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \\ X_p = l_{p1}F_1 + l_{p2}F_2 + \dots + l_{pm}F_m + e_p \end{cases}$$

donde:  $F_1, F_2, \dots, F_m$  son factores comunes;  $e_1, e_2, \dots, e_p$  son factores únicos o específicos;  $l_{jh}$  es el peso del factor  $h$  en la variable  $j$  y se denominan cargas factoriales. Cada una de las  $p$  variables observables es una combinación lineal de  $m$  factores comunes ( $m < p$ ) y de un factor único. De esta manera, todas las variables originales vienen influidas por todos los factores comunes, mientras que existe un factor único que es específico para cada variable. Tanto los factores comunes como los únicos no son observables.

#### 3.1.1.2 Matriz de correlación reproducida

En el análisis factorial se parte del supuesto de que las variables originales están correlacionadas entre sí. La matriz de correlación muestral refleja la correlación directa existente entre cada par de variables, ya que comparten unos mismos factores comunes (Uriel y Aldás, 2005). A la matriz formada por correlaciones poblacionales estimadas se le llama matriz de correlación reproducida. Si el modelo factorial es adecuado a los datos, la diferencia para cada par de variables entre el coeficiente de correlación muestral y el coeficiente de correlación reproducido será muy pequeña, ya que este último coeficiente mide la correlación entre dos variables a través de los respectivos coeficientes de correlación con los factores (Uriel y Aldás, 2005).

#### 3.1.1.3. Métodos

Existen varios métodos para la extracción de factores: componentes principales, mínimos cuadrados no ponderados (ULS-Unweighted Least Squares), mínimos cuadrados generalizados (GLS-Generalized Least Squares), y máxima verosimilitud (ML-Maximum Likelihood). En los métodos GLS y ML es necesario que se cumpla el supuesto de normalidad multivariante.

#### 3.1.1.4. Procedimiento

Choi *et al.* (2004) sugirieron una serie de pasos para refinar una escala por medio del AFE:

1. Llevar a cabo un AFE y ver si surge la estructura de factores hipotetizada.
2. Eliminar los ítems que están poco relacionados con sus factores hipotéticos o los que se asocian con más de un solo factor.
3. Usar los estimados de alfa de Cronbach y las correlaciones de ítem al total (o las cargas factoriales), para comprobar las fiabilidades de los ítems que miden cada factor de la hipótesis y eliminar los ítems que sean poco fiables.
4. Repetir los pasos 1 a 3, hasta que surja una estructura factorial clara.



### 3.1.1.5. Contrastes de validación del modelo

Los contrastes y estadísticos del AFE, se agrupan en dos bloques (Uriel y Aldás, 2005):

1. Los que se aplican previamente a la extracción de factores, cuyo objetivo es analizar la pertinencia de aplicación del AFE a un conjunto de variables. Los más usados son:

*Contraste de esfericidad de Barlett:* prueba si las variables originales están correlacionadas entre sí; si no lo estuvieran, no existirían factores comunes y no tendría sentido aplicar el AFE.

*Medidas de adecuación muestral:* la medida KMO (Kaiser-Meyer-Olkin) mide de adecuación de la muestra al AFE, la cual debe ser próxima a 1.

2. Los que se aplican *ex post* y que pretenden evaluar el modelo factorial estimado; en este tipo se enmarca la *medida de bondad de ajuste*, basado en un contraste de bondad de ajuste con una distribución Chi-cuadrada ( $\chi^2$ ), cuando el método de estimación aplicado ha sido el ML o el GLS.

### 3.1.1.6. Rotación de factores

Como afirman Uriel y Aldás (2005), la rotación de factores se ha ideado para obtener, a partir de la solución final, factores que sean fácilmente interpretables. Existen dos formas de realizar la rotación: la ortogonal y la oblicua. En la ortogonal, los ejes se rotan para que se preserve la incorrelación entre los factores (los ejes rotados son perpendiculares). El método más conocido es el Varimax. En la oblicua, los ejes no son perpendiculares, consiguiéndose una asociación más nítida de cada una de las variables con el factor correspondiente. El método más conocido es el Oblimin.

### 3.1.2. Análisis Factorial Confirmatorio (AFC)

El AFC provee una rigurosa evaluación de la dimensionalidad, proveyendo consistencia interna y externa.

#### 3.1.2.1. Identificación del modelo en el AFC

Uriel y Aldás (2005) sostienen que en el método AFC se dispone de una serie de datos (las varianzas-covarianzas muestrales de las variables observadas) y con ellos se ha de estimar una serie de parámetros (cargas factoriales, varianzas-covarianzas de los factores comunes y varianzas-covarianzas de los factores específicos o errores). Al igual que ocurre en un sistema de ecuaciones lineales, si no se imponen restricciones a los parámetros a estimar, habrá un número infinito de soluciones posibles. La identificación hace referencia a la cuestión de si los parámetros del modelo pueden o no ser determinados de forma única. Siguiendo a Ullman (1996), el investigador debería centrarse en las siguientes tareas:

1. Comparar el número de datos con el número de parámetros que han de estimarse.
2. Establecer una escala para los factores comunes.
3. Asegurar la identificabilidad de la parte del modelo que contiene la relación entre las variables observadas y los factores.
4. Fijar arbitrariamente el coeficiente de regresión del término de error al valor 1.

#### 3.1.2.2. Estimación de modelos en el AFC

Según Sharma (1996), el proceso de estimación del AFC sigue los siguientes pasos:

1. Dada la matriz de varianzas-covarianzas muestrales, se estiman los parámetros del modelo factorial hipotetizado.
2. Se determina el ajuste del modelo hipotetizado. En qué medida la matriz de varianzas-covarianzas estimada está próxima a la matriz de varianzas-covarianzas muestral.

Los procedimientos de estimación más comunes son: ULS (cuya ventaja reside en que no es necesaria la suposición de normalidad de las variables observadas), GLS, ML, libre distribución asintótica (ADF - Asymptotically Distribution Free) o mínimos cuadrados ponderados (WLS - Weighed Least Squares).

### 3.1.2.3. Bondad de ajuste del modelo estimado

Es necesario determinar hasta qué punto el modelo asumido se ajusta a los datos muestrales. Si se detectan problemas de ajuste, será necesario reespecificar el modelo. Se usan para ello los siguientes estadísticos.

Primero se tiene el estadístico  $\chi^2$  para el contraste global del modelo. Los supuestos que requiere son: la hipótesis de normalidad de las variables observadas y un tamaño muestral lo suficientemente grande para justificar las propiedades asintóticas del contraste. Como estos requisitos se cumplen rara vez, Long (1983) y Ullman (1996) sugieren que el estadístico debe aplicarse con precaución con muestras grandes, dado que incluso pequeñas diferencias entre las matrices de covarianza muestral y la estimada serán evaluadas como significativas. El desarrollo de los estadísticos *ad hoc* de bondad de ajuste (véase la Tabla 1) surgen ante la limitación anterior.

Los índices comparativos de ajuste comparan el valor del modelo teórico que se evalúa, con el del modelo independiente (consistente en uno sin ninguna relación entre los constructos del modelo); los índices de proporción de varianza miden la proporción de ajuste entre la matriz de varianzas-covarianzas poblacional y la muestral; los índices de grado de parsimonia indican cuán complejas son las relaciones del modelo con el independiente (un valor más próximo al modelo independiente indicará que el modelo es más simple que otro); y los índices basados en residuos indican la diferencia que existe entre la matriz de varianzas-covarianzas muestral y la poblacional (Uriel y Aldás, 2005).

Tabla 1. Índices *ad hoc* de bondad de ajuste para el AFC

Grupo	Índice	Siglas	Ajuste aceptable
Índices comparativos de ajuste	Normed Fit Index	NFI	0,90 – 0,95
	Nonnormed Fit Index	NNFI	0,95 – 0,97
	Comparative Fit Index	CFI	0,95 – 0,97
	Incremental Fit Index	IFI	> 0,90
Índices de proporción de varianza	Goodness of Fit Index	GFI	0,90 - 0,95
	Adjusted Goodness of Fit Index	AGFI	0,85 - 0,90
Índices de grado de parsimonia	Akaike Information Criterion	AIC	El menor en la comparación
	Consistent AIC	CAIC	El menor en la comparación
Índices basados en residuos	Root Mean Squared Residual	RMR	0,05 – 0,10
	RMR Error of Approximation	RMSEA	0,05 – 0,08

Fuente: adaptado de García (2011) y Uriel y Aldás (2005)

#### *3.1.2.4. Interpretación del modelo*

Se examina si los estimadores de los parámetros son razonables en dos sentidos (Uriel y Aldás, 2005): 1) ¿Toman valores adecuados teóricamente?; y 2) ¿son significativos? Para que exista compatibilidad de las estimaciones de los parámetros con el modelo estadístico, las respuestas a las siguientes preguntas deben ser negativas: a) ¿Existen correlaciones superiores a la unidad?; b) ¿existen cargas factoriales estandarizadas fuera del intervalo  $-1, 1$ ?; c) ¿los residuos estandarizados son anormalmente grandes?; y d) ¿hay estimaciones negativas de las varianzas? Si no es el caso, existiría una incompatibilidad debida a que, según Long (1983): 1) el modelo estaría mal especificado; 2) los datos no respaldarían la hipótesis de normalidad multivariante de las variables observadas; 3) la muestra sería demasiado pequeña; o 4) el modelo estaría cerca de no estar identificado, dificultando la estimación de algunos parámetros.

#### *3.1.2.5. Reespecificación del modelo*

Ullman (1996) señala que existen dos motivos para reespecificar un modelo (es decir, eliminar o introducir relaciones entre las variables que los conforman). Estos motivos son: 1) mejorar su ajuste; y 2) contrastar alguna hipótesis teórica.

### **3.2. Modelos de Estructuras de Covarianza (MEC)**

#### *3.2.1. Propósito*

El MEC permite examinar simultáneamente una serie de relaciones de dependencia y es útil cuando una variable dependiente se convierte en variable independiente en ulteriores relaciones de dependencia (García, 2011; Poza y Fernández, 2011). Además, muchas de las mismas variables afectan a cada una de las variables dependientes, pero con efectos distintos (Cupani, 2012).

El MEC es una técnica estadística multivariante que puede ser entendida como una extensión de la regresión múltiple y el análisis factorial (Poza y Fernández, 2011). Posee algunas características particulares que lo diferencian de otras técnicas multivariadas (Kahn, 2006). Una de las diferencias es la capacidad de estimar y evaluar la relación entre constructos no observables, denominadas variables latentes, de forma múltiple y cruzada, que solo pueden ser medidas mediante variables observables, y tener en cuenta el error de medida en el proceso de estimación (Poza y Fernández, 2011).

Otra característica es que para interpretar los resultados del MEC se deben evaluar varias pruebas estadísticas e índices que determinan si la estructura teórica propuesta presenta un buen ajuste a los datos empíricos (García, 2011). Este ajuste se verifica si los valores de los parámetros estimados reproducen tan estrechamente como sea posible la matriz observada de covarianza (Kahn, 2006).

Como indican Uriel y Aldás (2005), los modelos de estructuras de covarianza (MEC), se descomponen en dos componentes: 1) la componente estructural, que incluye las relaciones entre los factores latentes; y 2) la componente de medida, que incluye las variables observadas que miden los factores latentes. Según Cupani (2012) el modelo estructural es el modelo guía, en el cual la teoría y no la experiencia, permitirá al investigador distinguir a las variables independientes que predicen cada

variable dependiente, mientras que el objetivo del modelo de medida es corroborar la idoneidad de los indicadores seleccionados en la medición de los constructos de interés.

### 3.2.2. Pasos para la modelización

Los especialistas en el MEC (García, 2011; Uriel y Aldás, 2005) consideran seis pasos a seguir para aplicar esta técnica: especificación, identificación, estimación de parámetros, evaluación del ajuste, reespecificación del modelo e interpretación de resultados (Kline, 2005). Estos pasos son similares a los descritos en la técnica AFC, aunque tendrán mayor complejidad debido a que el número de parámetros suele ser muy superior.

## 4. ESTUDIO EMPÍRICO PARA EL DISEÑO DE INSTRUMENTOS DE MEDIDA

### 4.1. Contexto

Se realizó un estudio empírico en la Universidad Privada Boliviana (UPB), con el objetivo de obtener instrumentos válidos y fiables de los constructos de la calidad de servicio de docencia, el valor percibido del servicio de docencia, la satisfacción estudiantil y las intenciones de comportamiento de postdocencia de los estudiantes.

### 4.2. Estudio exploratorio

A fin de obtener instrumentos que tengan validez de contenido, en primer lugar, se llevó a cabo un estudio exploratorio, siguiendo las fases sugeridas por Martínez (1995): 1) identificar el propósito; 2) especificar *a priori* las restricciones de operación: tiempo, medios y situaciones; 3) identificar el dominio del constructo; 4) especificar el formato de los ítems; 5) preparar un conjunto de ítems que cubran la totalidad del dominio; 6) un plan para seleccionar ítems relevantes; y 7) configurar el diseño definitivo. Para realizar las fases 1 a 4, se definieron aspectos importantes en la construcción del test, clasificados en:

#### 1) Factores contextuales externos:

- Propósito del instrumento: diagnóstico y predictivo.
- Tiempo de aplicación: sin restricciones; aproximadamente 20 minutos.
- Escenario temporal de aplicación: últimos días de impartición de una materia.

#### 2) Factores contextuales internos:

- Definición del dominio del constructo: la calidad de servicio de docencia presenta la calidad técnica y la funcional, donde se percibe la calidad cuando el docente excede las expectativas estudiantiles. La deficiencia es medida directamente (Snipes *et al.*, 2006).
- Tipo de formato: se miden percepciones o actitudes estudiantiles. Las respuestas serán tipo Likert, con una escala de 7 puntos, con (1) “mucho peor de lo esperado” y (7) “mucho mejor de lo esperado”.

Para realizar las fases 5 y 6 se utilizaron como base las siguientes escalas: 1) generales, que miden la calidad de servicio en la Educación Superior (LeBlanc y Nguyen, 1997; Abdullah, 2006) y en

otros servicios (Parasuraman *et al.*, 1991; Dabholkar *et al.*, 1996; Markovic, 2006); y 2) específicas, que miden la calidad de servicio de docencia (Banwet y Datta, 2003; Snipes *et al.*, 2006).

A partir de la revisión de instrumentos en la literatura, se diseñó un cuestionario estructurado, que contenía los siguientes aspectos, dividido en dos partes: 1) la Parte I tenía el propósito de realizar una recopilación de los principales aspectos de interés de alumnos y docentes acerca de cómo medir la calidad de servicio percibida de docencia, mediante la técnica de incidentes críticos y una encuesta de los principales aspectos para su medición; y 2) la Parte II pedía a los encuestados que analizaran un instrumento desarrollado por expertos del medio educativo en cuanto a: ítems y dimensiones relevantes, pertinencia de las directrices, confusiones al llenar el cuestionario, preguntas de doble sentido, etc.

Se tomó como muestra de juicio a estudiantes elegidos por facultad y carreras y con excelencia académica; y otra muestra de conveniencia a docentes y jefes académicos, tomando en cuenta su experiencia educativa. Se encuestó a 100 miembros de la UPB, obteniendo una tasa de respuesta del 40%. Tomando en cuenta las opiniones de los encuestados se diseñó una batería con ítems relevantes para la medición de la calidad de servicio de docencia.

Para realizar la fase 7, se llevó a cabo una aplicación piloto. Se administró la batería del instrumento a un curso elegido al azar, obteniéndose 30 encuestas. Las deficiencias del llenado del cuestionario condujeron a varias acciones correctivas. De esa manera, se obtuvo el diseño definitivo de una batería para la medición de la calidad de servicio de docencia percibida por los estudiantes, que constaba de 40 ítems, para encarar el estudio de validez y fiabilidad. También se generaron instrumentos para la medición de la calidad de servicio global de docencia (CSG), valor percibido del servicio de docencia (VAL), satisfacción estudiantil de la docencia (SAT), e intenciones de comportamiento de postdocencia (IC). En todos estos instrumentos se usó una escala de 7 puntos tipo Likert, con (1) “totalmente en desacuerdo” y (7) “totalmente de acuerdo”.

#### **4.3. Obtención de instrumentos válidos y fiables de la calidad de servicio de docencia y de sus constructos relacionados**

El instrumento de 40 ítems obtenido en el estudio exploratorio fue sometido luego a un AFE y a un AFC, para encontrar su dimensionalidad adecuada, y se realizaron pruebas de validez de constructo y fiabilidad.

##### *4.3.1. Recopilación de datos*

La población fue de 1145 alumnos de pregrado matriculados en la UPB, en el primer semestre del 2011. Para la recopilación de datos se usó el método del cuestionario y un muestreo de oportunidad, utilizando la logística que tiene esta universidad para la evaluación del desempeño docente. El cuestionario fue repartido por los jefes de carrera, los últimos días antes que culmine una materia, cuidando la confidencialidad de la información.

Se obtuvieron 600 encuestas, que fueron divididas en dos muestras, 300 para el AFE y 300 para el AFC y el MEC. Al respecto, Martínez-López *et al.* (2004) indican que el tamaño muestral

puede obtenerse de una relación de 5 individuos por parámetro que se quiere estimar en el modelo. Para el AFE se tienen 40 ítems, para el AFC 13 ítems y para el MEC 30, por lo cual el tamaño muestral elegido (300) cumple la regla de proporción 5:1.

#### 4.3.2. Idoneidad de los datos

Siguiendo a Uriel y Aldás (2005), las pruebas de idoneidad se dividieron en dos: a) previas a la técnica, para comprobar la confiabilidad de los datos; y b) propias a la técnica, para comprobar hipótesis subyacentes, como la normalidad y linealidad.

En el primer caso, se eliminaron los cuestionarios mal llenados o con patrones anómalos. En el segundo caso se verificó mediante el método del gráfico Chi-cuadrado (Johnson y Wichern, 1998) que no se cumplía el supuesto de normalidad multivariante. Por ese motivo, se usaron el método de componentes principales para realizar el AFE, y el método ULS, para realizar el AFC y el MEC. También se comprobó que las relaciones entre los ítems eran aproximadamente lineales.

#### 4.3.3. Análisis Factorial Exploratorio (AFE)

Con la primera muestra de tamaño 300, se procedió a realizar el AFE. Las medidas de adecuación muestral para la calidad de servicio percibida de docencia son mostradas en la Tabla 2, las cuales han sido obtenidas usando el programa SPSS 18.0.

Tabla 2. Medidas de adecuación muestral de la calidad de servicio de docencia

Medida de adecuación muestral de Kaiser-Meyer-Olkin		0,961
Prueba de esfericidad de Bartlett	Chi-cuadrado aproximado	9052,227
	Gl.	780
	Sig.	0,000

La prueba de esfericidad de Bartlett es significativa, así que existe relación entre los ítems de la calidad de servicio, aspecto primordial para el AFE. Como la medida KMO es próxima a 1, se pudo concluir que los datos son adecuados para realizar el análisis.

Los criterios para la eliminación de ítems fueron los siguientes (Ladhari, 2008):

- Puntajes de carga de menos de  $\pm 0,45$  (Markovic, 2006).
- Cargas cruzadas de más de  $\pm 0,40$  (Janda *et al.*, 2002).

Después de la eliminación de ítems, se realizó un análisis de fiabilidad, para ver si la eliminación de algunos mejoraba el coeficiente alfa de Cronbach para cada dimensión. El análisis factorial con el método de componentes principales (rotación varimax) dio como resultado 6 dimensiones (véase la Tabla 3).

Por su parte, la Tabla 4 resume los pasos seguidos para la eliminación de ítems.

En las tres primeras fases se desecharon ítems que cargaron en dos dimensiones o débilmente. En la cuarta eliminación se desecharon ítems para mejorar el alfa de Cronbach. Se logró una varianza acumulada extraída de 71%. Se observa que la medida KMO y el índice de Bartlett después de las eliminaciones tuvieron valores adecuados.

La estructura factorial resultante se muestra en la Tabla 5.

Tabla 3. AFE para la calidad de servicio de docencia

Ítem	Componente						Ítem	Componente					
	1	2	3	4	5	6		1	2	3	4	5	6
CS1					0,797		CS21		0,763				
CS2			0,455		0,404		CS22	0,475	0,460				
CS3					0,815		CS23		0,627				
CS4					0,561		CS24		0,609				
CS5						0,616	CS25		0,690				
CS6			0,689				CS26		0,564				
CS7			0,684				CS27	0,435	0,438	0,510			
CS8			0,620				CS28			0,678			
CS9			0,462	0,514			CS29	0,446		0,524			
CS10						0,507	CS30	0,549		0,464			
CS11	0,432			0,482			CS31	0,463	0,425				
CS12				0,456			CS32	0,556					
CS13		0,416	0,543				CS33			0,579			
CS14			0,479	0,402			CS34	0,494					
CS15				0,739			CS35	0,766					
CS16		0,437		0,547			CS36	0,692					
CS17		0,453		0,444			CS37	0,676					
CS18		0,630					CS38	0,706					
CS19		0,487	0,417				CS39	0,741					
CS20		0,777					CS40	0,702					

Tabla 4. Procedimiento para lograr una estructura factorial adecuada para la calidad de servicio de docencia

Características	AFE inicial	1ª eliminación	2ª eliminación	3ª eliminación	4ª eliminación
Ítems eliminados		2, 9, 11, 13, 14,16, 17, 19,22, 27, 29, 30, 31	4, 7, 8, 10, 12, 26, 28, 32, 37	5, 6, 34	15, 33
Dimensiones	6	4	3	3	3
KMO	0,961	0,948	0,924	0,91	0,901
Prueba de Bartlett	9052,227 (g.l. 780)	5308,524 (g.l. 351)	3011,059 (g.l. 153)	2538,304 (g.l. 105)	2310,175 (g.l. 78)
Varianza acumulada extraída (%)	66,392	63,131	61,202	65,715	70,866

Tabla 5. Estructura factorial resultante de la calidad de servicio de docencia

Ítem	Componente		
	1	2	3
CS1			0,856
CS3			0,863
CS18	0,718		
CS20	0,830		
CS21	0,832		
CS23	0,728		
CS24	0,713		
CS25	0,781		
CS35		0,813	
CS36		0,741	
CS38		0,787	
CS39		0,836	
CS40		0,779	

Se observa una estructura clara, con 3 dimensiones (seguridad, resultado y tangibles), 13 ítems, y en las cuales todos los ítems cargaron en sus dimensiones hipotéticas. El instrumento para medir la calidad de servicio percibida de docencia se muestra en el Anexo A.1. El coeficiente de fiabilidad ponderado fue de 0,993, indicando que la batería multidimensional de la calidad de servicio de docencia es fiable (tiene consistencia interna). Usando el AFE se confirmó que la calidad de servicio global, valor percibido, satisfacción estudiantil e intenciones de postdocencia son unidimensionales. La Tabla 6 muestra las características de los cuatro constructos.

Tabla 6. AFE de la calidad de servicio global de comportamiento de postdocencia

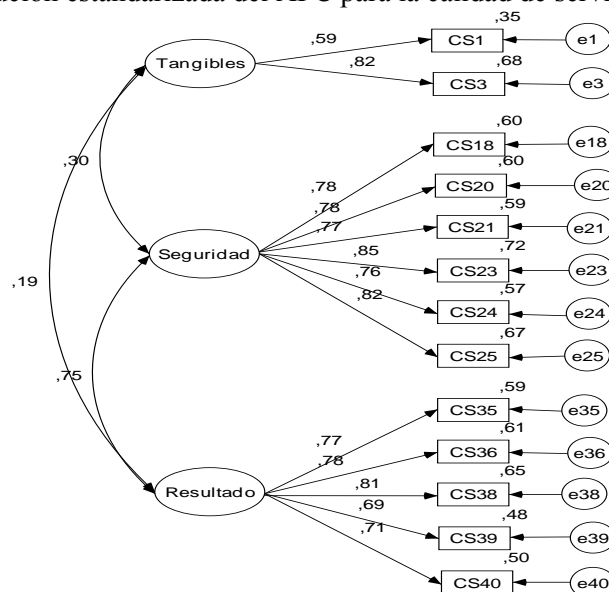
Características AFE	Calidad de servicio global (CSG)	Valor percibido (VAL)	Satisfacción (SAT)	Intenciones de comportamiento (IC)
Ítems eliminados	-	-	-	IC5
Dimensiones	1	1	1	1
Prueba KMO	0,866	0,837	0,842	0,894
Prueba de Bartlett	1262,379 (6 g.l.)	982,190 (6)	1455,922 (6)	1856,778 (10)
Varianza acumulada extraída (%)	87,689	81,441	89,531	87,167
Alfa de Cronbach	0,953	0,922	0,961	0,963

Todos los ítems se retuvieron en su respectivo constructo, excepto el ítem 5 de las intenciones de comportamiento (IC5), que fue eliminado: “Volvería a pasar clases en esta aula (incluye el equipo didáctico)”. De acuerdo a los valores de la prueba KMO, las muestras son adecuadas para el AFE. Por otro lado, la varianza porcentual extraída con la estructura unidimensional de los cuatro constructos considerados es adecuada y bastante alta. Los alfas de Cronbach ( $\alpha$ ) de los 4 constructos son superiores a 0,9. En el Anexo A.2 se exponen las baterías para la medición de estos constructos.

#### 4.3.4. Análisis Factorial Confirmatorio (AFC)

Se usó el programa AMOS 6.0 para generar los resultados del AFC. Se optó por el método ULS, ya que los datos no cumplieron el supuesto de normalidad multivariante. La Figura 3 muestra la solución estandarizada del AFC.

Figura 3. Solución estandarizada del AFC para la calidad de servicio de docencia





Se puede observar que todos los coeficientes de regresión estandarizados (cargas factoriales) entre los ítems y las dimensiones de la calidad de servicio tienen valores altos comprendidos entre 0 y 1 (como sugieren Uriel y Aldás, 2005) y todos son positivos, indicando una relación causal directa de las dimensiones hacia los ítems que los conforman. Esta fuerza de la relación de cada dimensión con sus ítems es un indicativo de una estructura factorial robusta.

Para determinar si el modelo asumido se ajustaba a los datos muestrales (es decir, presentaba un buen ajuste), se calcularon los índices ad hoc mostrados en la Tabla 7.

Tabla 7. Índices ad hoc para el AFC de la calidad de servicio de docencia

Grupo	Índice	Modelo independiente	Modelo analizado
Índices comparativos de ajuste	NFI		0,982
Índices de proporción de varianza	GFI	0,296	0,987
	AGFI	0,178	0,981
Índices basados en residuos	RMR	0,721	0,097

Los índices comparativos de ajuste y de proporción de varianza deben ser superiores a 0,9. Como se observa el NFI sobrepasa ese valor y los índices GFI y AGFI son cercanos a 1. Los índices basados en residuos deben ser menores a 0,10, y se observa que el RMR está dentro del rango permitido.

Por lo expuesto, se puede concluir que el modelo de la calidad de servicio de docencia presenta buenos índices de ajuste, lo que quiere decir que el AFC ha confirmado la estructura factorial formulada por el AFE.

#### 4.3.5. Análisis de validez de la calidad de servicio de docencia

La validez discriminante (grado al cual una medida no correlaciona fuertemente con otras medidas que se supone difieren entre sí) se confirma mediante las correlaciones entre las dimensiones de la calidad de servicio (véase la Tabla 8), ya que no existe un valor muy alto.

Tabla 8. Correlaciones entre las dimensiones de la calidad de servicio de docencia

Correlaciones	Tangibles	Seguridad	Resultado
Tangibles	1,000	0,238	0,146
Seguridad	0,238	1,000	0,658
Resultado	0,146	0,658	1,000

La validez convergente (es decir, el grado al cual un grupo de ítems que representan un constructo convergen sobre el constructo) se verifica observando las cargas factoriales de cada ítem en su dimensión (véase la Figura 3) y se comprueba que cada ítem ha cargado fuertemente en la dimensión hipotética.

La validez concurrente y predictiva se comprueban con las correlaciones entre constructos relacionados a la calidad de servicio (véase la Tabla 9).

Tabla 9. Matriz de correlaciones entre la calidad de servicio de docencia, valor percibido, satisfacción e intenciones de comportamiento de postdocencia

Correlaciones	CS	CSG	VAL	SAT	IC
CS	1,000	0,675	0,629	0,653	0,672
CSG	0,675	1,000	0,704	0,814	0,822
VAL	0,629	0,704	1,000	0,741	0,733
SAT	0,653	0,814	0,741	1,000	0,909
IC	0,672	0,822	0,733	0,909	1,000

La validez concurrente es el grado al cual los puntajes de la calidad de servicio, están asociados con medidas relacionadas conceptualmente, pero que son medidas al mismo tiempo, como la calidad de servicio global. La validez predictiva tiene el mismo concepto que la concurrente, pero las medidas relacionadas deben ser tomadas posteriormente, como ocurre con la satisfacción o las intenciones de comportamiento (Martínez, 1995).

El modelo tiene validez concurrente, ya que la calidad de servicio (CS) está relacionada con la calidad de servicio global (CSG); y tiene validez predictiva, porque está relacionada a constructos que son su consecuencia, como la satisfacción (SAT), el valor percibido (VAL) y las intenciones de comportamiento (IC).

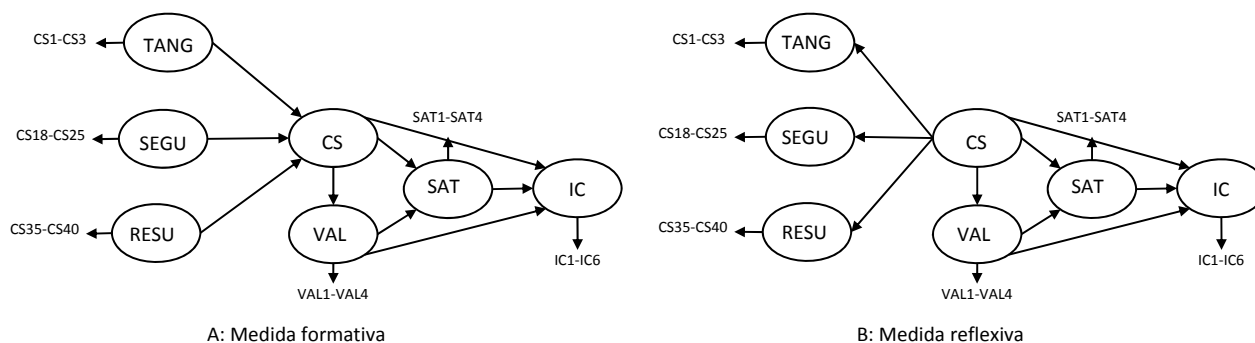
## 5. ANÁLISIS DE COMPARACIÓN DE LOS MODELOS ESTRUCTURALES DE EVALUACIÓN DEL SERVICIO DE DOCENCIA

### 5.1. Caracterización de los modelos reflexivo, formativo y de antecedentes

El modelo estructural que se analizó en esta investigación es denominado “comprehensivo”, similar al que plantearon Brady *et al.* (2005) y Choi *et al.* (2004), adaptado al contexto del servicio de docencia en la Educación Superior. Es comprehensivo en el sentido de que la calidad de servicio, el valor y la satisfacción influyen directa y conjuntamente a las intenciones de comportamiento.

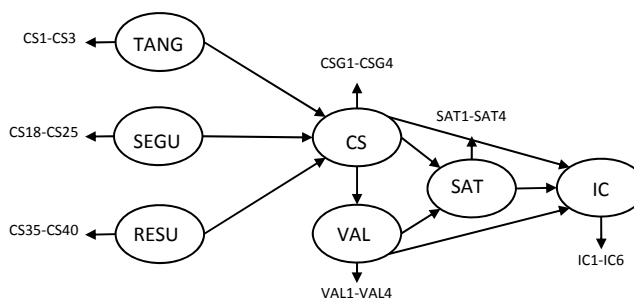
Los modelos reflexivo y formativo de la calidad de servicio de docencia, que se analizaron en esta investigación, se muestran en la Figura 4. En el panel A, las dimensiones de tangibles, seguridad y resultado hacen una contribución única al constructo de la calidad de servicio; es decir, causan el constructo. En el panel B, la calidad de servicio causa las dimensiones, siendo éstas intercambiables entre sí. En ambos modelos la calidad de servicio está conceptualizada como un constructo multidimensional de segundo orden.

Figura 4. Modelos reflexivo y formativo de la calidad de servicio de docencia



El modelo de antecedentes es el mostrado en la Figura 5. Diverge del modelo formativo en que las dimensiones de la calidad de servicio de docencia (tangibles, seguridad y resultado) son constructos antecedentes y la calidad de servicio de docencia se mide haciendo uso de una batería unidimensional compuesta de 4 ítems (CSG1 – CSG4).

Figura 5. Modelo de antecedentes de la calidad de servicio de docencia



Los modelos expuestos en las Figuras 4 y 5 han sido simplificados a fin de mostrar las relaciones más importantes. Toda variable no observable del modelo (componente de medida) ha sido medida mediante indicadores reflexivos, mediante los instrumentos mostrados en los Anexos A.1 y A.2.

## 5.2. Comparación de los modelos formativo, reflexivo y de antecedentes

### 5.2.1. Bases de comparación

Los tres modelos presentan una especificación adecuada, según el examen propuesto por Uriel y Aldás (2005): no han mostrado correlaciones superiores a la unidad, sus cargas factoriales estandarizadas no sobrepasan la unidad, no existen residuos extremadamente grandes y no se manifestaron estimaciones negativas de las varianzas.

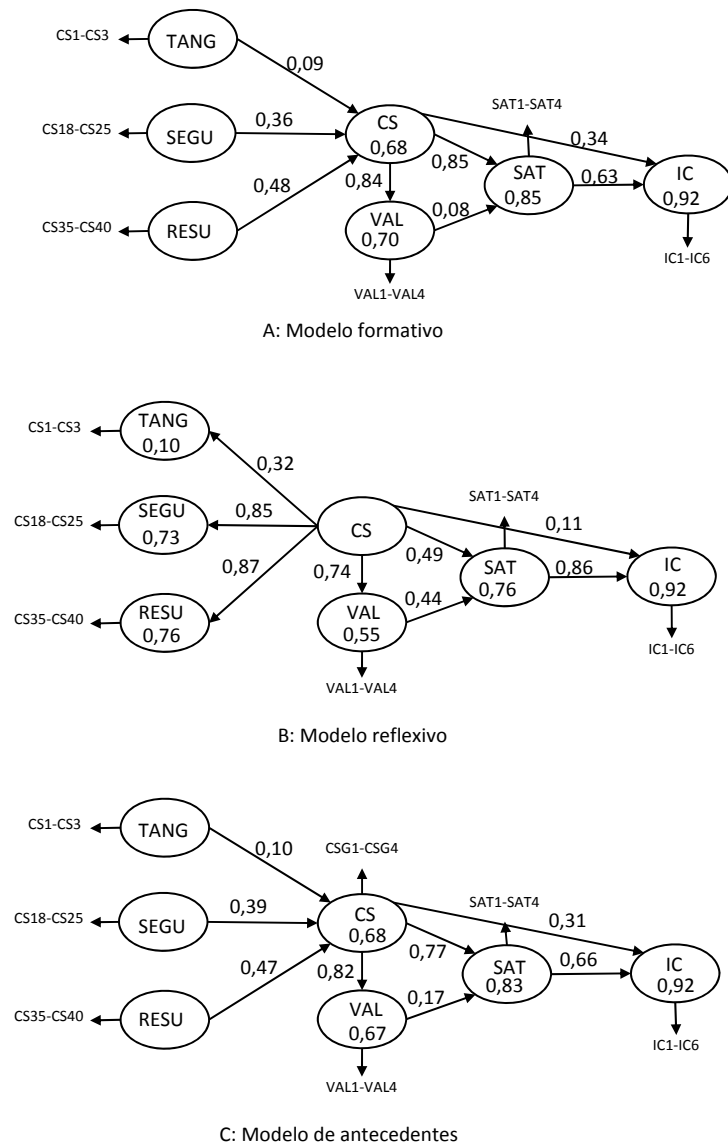
Por lo tanto, como los tres modelos están bien especificados, una de las maneras de probar la superioridad de los tres modelos es la de comparar de medidas de bondad de ajuste *ad hoc*. El modelo que presente mejores medidas de bondad de ajuste será el modelo más adecuado de medición de la calidad de servicio.

Otro criterio es la comparación de las correlaciones múltiples cuadradas entre los tres modelos. Bollen (1989) sugiere que si un modelo tiene correlaciones múltiples cuadradas de sus variables mayores al de otro modelo, entonces tiene mejor especificación de modelo de medición, ya que presenta mejor poder predictivo.

### 5.2.2. Diferencias entre los modelos formativo, reflexivo y de antecedentes

Primero se realizó una comparación entre las cargas factoriales y las relaciones causales entre los tres modelos, para determinar sus diferencias (véase la Figura 6). Estos valores fueron determinados mediante el MEC usando el programa AMOS 6.0.

Figura 6. Resultados estandarizados (MEC) de los modelos formativo, reflexivo y de antecedentes



Los valores encerrados en los círculos debajo de cada constructo son las correlaciones múltiples cuadradas ( $R^2$ ). En los tres modelos, la ruta del valor percibido con las intenciones de comportamiento (VAL-IC) no fue significativa (mostró una carga factorial estandarizada entre 0,04 y 0,05, dependiendo del modelo analizado); por lo tanto fue eliminada.

Se puede observar que existen diferencias en las relaciones causales entre la calidad de servicio, sus dimensiones y los constructos a los cuales está relacionada, entre uno y otro modelo. Estas diferencias establecen que la elección adecuada del modelo de medida es de fundamental importancia. Las principales conclusiones que se puede extraer de la comparación de los tres modelos son las siguientes:

- El modelo reflexivo favorece las relaciones valor-satisfacción y satisfacción-intenciones de comportamiento. Todas las relaciones de la calidad de servicio se ven disminuidas.

- El modelo formativo favorece más la relación calidad de servicio-valor, calidad de servicio-satisfacción y calidad de servicio-intenciones de comportamiento. En otras palabras, intensifica la relación de la calidad de servicio con sus constructos relacionados.
- El modelo de antecedentes intensifica aún más que el modelo formativo las relaciones de la calidad de servicio con sus constructos relacionados.
- En los tres modelos la satisfacción se constituye en un mediador poderoso entre la calidad de servicio y el valor, sobre las intenciones de comportamiento.

### 5.2.3. Verificación de la superioridad entre los tres modelos

En la Tabla 10 se comparan los índices de bondad de ajuste de los tres modelos. Como se puede apreciar, existe un mejor ajuste del modelo de antecedentes, comparado con el reflexivo y el formativo, aunque las diferencias porcentuales son del orden del 0,1% para los grupos de índices comparativos de ajuste y de proporción de la varianza, y del orden del -8% para el grupo de índices basados en los residuos. Prácticamente los índices de ajuste del modelo formativo son similares a los del modelo reflexivo.

Tabla 10. Comparación de índices de bondad de ajuste entre los tres modelos

Índices de ajuste		Modelo		
Grupo	Siglas	Formativo	Reflexivo	Antecedentes
Chi-cuadrado	$\chi^2$	879,785 (288 g.l.)	900,987 (291 g.l.)	992,879 (394 g.l.)
Índices comparativos de ajuste	NFI	0,992	0,992	0,994
Índices de proporción de varianza	GFI	0,993	0,993	0,994
	AGFI	0,992	0,992	0,993
Índices basados en residuos	RMR	0,092	0,093	0,085

El índice RMR (Root Mean Squared Residual) para el modelo de antecedentes es menor que para los otros dos modelos, indicando que con el modelo de antecedentes se ha logrado una mayor similitud entre la matriz de varianzas-covarianzas muestral y la poblacional.

En el grupo de índices comparativos de ajuste (que comparan el valor del modelo teórico que se evalúa, con el del modelo independiente), el NFI posee un valor mayor en el modelo de antecedentes que en los otros dos modelos, sugiriendo que el modelo de antecedentes está más alejado de la semejanza con el modelo independiente que los otros dos modelos.

En el grupo llamado “índices de proporción de la varianza” (que establecen la proporción entre la matriz de varianzas-covarianzas poblacional y la muestral), el GFI y AGFI presentan mayores valores en el modelo de antecedentes que en los otros dos modelos, indicando que el modelo de antecedentes explica una proporción mayor de la varianza poblacional que los otros dos modelos. Estos resultados indican que un modelo de antecedentes se desempeña mejor que el reflexivo o el formativo (el modelo presenta una mejor especificación de medición), probando así su superioridad.

Adicionalmente, se puede observar (véase la Figura 6) que los predictores del constructo de las intenciones de comportamiento de postdocencia de los tres modelos (antecedentes, reflexivo y formativo) explican una proporción igual de su varianza, indicando que todos los modelos

considerados tienen buen poder predictivo. La correlación múltiple cuadrada de las intenciones de comportamiento de postdocencia es del 92%, indicando que los predictores de ese constructo, que son la calidad de servicio de docencia, el valor percibido y la satisfacción estudiantil explican el 92% de la variabilidad de las intenciones de comportamiento de postdocencia.

## **6. RESULTADOS**

Mediante un análisis de estructuras de covarianza (MEC) se comprobó, a través de índices de bondad de ajuste y correlaciones múltiples cuadradas, que el modelo de antecedentes (unidimensional) para las dimensiones de la calidad de servicio percibida de docencia, es más idóneo que el modelo con enfoque de medición reflexivo y formativo. Por otro lado, se probó que el modelo de antecedentes intensifica más que el formativo y el reflexivo las relaciones calidad de servicio-valor, calidad de servicio-satisfacción y calidad de servicio-intenciones de comportamiento. Los índices de bondad de ajuste del modelo de antecedentes fueron bastante altos en comparación con los de los otros dos modelos.

Con respecto a la relación entre constructos en el modelo estructural estudiado, se comprobó el rol mediador significativo de la satisfacción entre la calidad de servicio percibida de docencia y el valor sobre las intenciones de comportamiento de postdocencia. La relación entre el valor percibido y las intenciones de comportamiento no fue significativa.

## **7. DISCUSIÓN DE RESULTADOS**

El resultado encontrado en esta investigación, “el modelo de antecedentes de la calidad de servicio de docencia tiene mejor modelo de especificación de medición que el modelo formativo o reflexivo, dentro de un modelo estructural de evaluación del servicio de docencia”, es congruente con los hallazgos realizados en otros escenarios de servicios por Dabholkar *et al.* (2000) y Wang y Lo (2002) y, en la Educación Superior, por Banwet y Datta (2003).

La mayoría de las investigaciones han elegido el enfoque reflexivo sin indagar si es la especificación de medición más idónea para la calidad de servicio (Ladhari, 2009; Alén y Fraiz, 2006). Varios investigadores (Podsakoff *et al.*, 2006; Howell *et al.*, 2007) han sugerido que la mejor especificación de medición para un constructo multidimensional, como es la calidad de servicio, es formativa, debido a que sus indicadores están conceptualizados como determinantes exógenos del constructo, no son necesariamente intercambiables, no necesariamente covarían entre sí en un nivel alto y no se espera que tengan redes nomológicas similares. Sin embargo, en esta investigación se ha constatado que el modelo formativo presenta prácticamente los mismos índices de ajuste que los del modelo reflexivo.

## **8. IMPLICACIONES EN LA GESTIÓN**

En primer lugar, la asignación de recursos de marketing varía cuando se aplican diferentes métodos de conceptualización. Bajo el modelo formativo, los recursos probablemente se destinarán a aumentar las habilidades de los estudiantes (resultado del servicio) más que para mejorar la adquisición de seguridad en el proceso de adquisición de esas habilidades y no asignar recursos a la mejora de los

elementos tangibles en el aula. Bajo el modelo unidimensional (de antecedentes) existe casi la misma importancia entre seguridad y resultado en la percepción de la calidad de docencia, mientras que todos los componentes se tratan como críticos en el modelo reflexivo. Por lo tanto, la percepción de la calidad de servicio de los consumidores no podrá ser mejorada eficazmente si se aplican modelos errados. Tomando en cuenta el modelo de antecedentes, para mejorar la calidad de servicio de docencia, se deberá medir sus antecedentes como constructos separados (tangibles, seguridad y resultado) y la calidad de servicio con la especificación unidimensional multi-ítem.

En segundo lugar, las fuerzas de las relaciones estructurales varían de acuerdo al modelo elegido. Si se adopta el enfoque de antecedentes, los programas de mejora deberán estar enfocados en lograr mayor calidad de servicio de docencia; si se adoptan los enfoques formativo o reflexivo, se deberán enfocar en lograr mayor satisfacción estudiantil, de modo que se obtengan intenciones de comportamiento positivas.

En ninguno de los modelos el valor tiene efecto directo sobre las intenciones de comportamiento de postdocencia, de modo que su medición puede resultar innecesaria para el mejoramiento global del servicio de docencia.

## **9. LIMITACIONES E INVESTIGACIONES FUTURAS**

A pesar de la parsimonia del modelo de antecedentes, es necesario seguir investigando la posibilidad de generalizar el modelo propuesto, llevando a cabo estudios en otras universidades e institutos de educación superior. Específicamente, se deben comprobar tres aspectos importantes: a) la superioridad del modelo de antecedentes frente a los modelos formativo y reflexivo de evaluación del servicio de docencia; b) la mínima contribución de los elementos tangibles para mejorar la calidad de servicio; y c) el hecho que el valor percibido del servicio de docencia no contribuya de ninguna manera a generar intenciones de comportamiento de postdocencia en los estudiantes.

Por otro lado, se debe tener en cuenta que la fuerza de las relaciones del modelo pueden ser susceptibles de cambiar con el tiempo, por lo cual es necesario realizar estudios de carácter longitudinal.

Por último, podría ser interesante incluir en el modelo a los antecedentes del valor percibido del servicio de docencia (valor funcional, epistémico, del dinero, condicional, social, etc.) (Lin *et al.*, 2005) y analizar si éstos contribuyen a incrementar el efecto del valor percibido sobre las intenciones de comportamiento de postdocencia.

## **REFERENCIAS**

- Abdullah, F. (2006): "Measuring service quality in higher education: HEdPERF versus SERVPERF", *Marketing Intelligence & Planning*, 24 (1): 31–47.
- Alén, M.E.; Fraiz, J.A. (2006): "Evaluación de la relación existente entre la calidad de servicio, la satisfacción y las intenciones de comportamiento en el ámbito del turismo termal", *Revista Europea de Dirección y Economía de la Empresa*, 15 (3): 171–184.

- Asubonteng, P.; McCleary, K.J.; Swan, J.E. (1996): "SERVQUAL revisited: a critical review of service quality", *Journal of Service Marketing*, 10 (6): 62–81.
- Banwet, D.K.; Datta, B. (2003): "A study of the effect of perceived lecture quality on post-lecture intentions", *Work Study*, 52 (5): 234–243.
- Bollen, K. (1989): "Structural equations with latent variables", John Wiley & Sons, New York, NY.
- Boulding, W.; Kalra, A.; Staelin, R.; Zeithaml, V.A. (1993): "A dynamic process model of service quality: from expectations to behavioral intentions", *Journal of Marketing Research*, 30: 7–27.
- Brady, M.K.; Cronin, J.J.; Brand, R.R. (2002): "Performance-only measurement of service quality: a replication and extension", *Journal of Business Research*, 55: 17–31.
- Brady, M.K.; Knight, G.A.; Cronin, J. Jr.; Tomas G.; Hult, M.; Keillor, B.D. (2005): "Removing the contextual lens: a multinational, multi-setting comparison of service evaluation models", *Journal of Retailing*, 81 (3): 215–230.
- Carman, J.M. (1990): "Consumer perceptions of service quality: an assessment of the SERVQUAL dimensions", *Journal of Retailing*, 66 (1): 33–55.
- Carman, J.M. (2000): "Patient perceptions of service quality: combining the dimensions", *Journal of Services Marketing*, 14 (4): 337–352.
- Choi, K.S.; Cho, W.H.; Lee S.; Lee, H.; Kim, C. (2004): "The relationships among quality, value, satisfaction and behavioral intention in health care provider choice: A South Korean study", *Journal of Business Research*, 57: 913–921.
- Churchill, G.A.Jr. (1979): "A paradigm for developing better measures of marketing constructs", *Journal of Marketing Research*, 16 (1): 64–73.
- Coltman, T.; Devinney, T.M.; Midgley, D.F.; Venai, S. (2008): "Formative versus reflective measurement models: Two applications of formative measurement", *Journal of Business Research*, 61(12): 1250–1262.
- Cronin, J.J.; Taylor, S.A. (1992): "Measuring Service Quality: a reexamination and extension", *Journal of Marketing*, 56: 55–68.
- Cupani, M. (2012): "Análisis de Ecuaciones Estructurales: conceptos, etapas de desarrollo y un ejemplo de aplicación", *Revista Tesis*, 1: 186–199.
- Dabholkar, P.A.; Sheperd, C.D.; Thorpe, D.I. (2000): "A comprehensive framework for service quality: an investigation of critical conceptual and measurement issues through a longitudinal study", *Journal of Retailing*, 76 (2): 139–173.
- Dabholkar, P.A.; Thorpe, D.I.; Rentz, J.O. (1996): "A measure of service quality for retail stores", *Journal of the Academy of Marketing Science*, 24: 3–16.
- Diamantopoulos, A. (2008): "Formative indicators: Introduction to the special issue", *Journal of Business Research*, 61(12): 1201–1202.
- Domínguez, M.; Blancas, F.J.; Guerrero, F.M.; González, M. (2011): "A Critical Review to Construct Composite Indicators", *Revista de Métodos Cuantitativos para la Economía y la Empresa*, 11: 41–70.



- Fedriani, E.M.; Martín, A.M. (2011): “A Multidimensional Poverty Indicator Based on Euclidean Geometry”, *Revista de Ciencias Sociales (RCS)*, XVII (4): 625–639.
- García, M.A. (2011): “Análisis causal con ecuaciones estructurales de la satisfacción ciudadana con los servicios municipales”. Proyecto Fin de Máster. Universidad de Santiago de Compostela.
- Grönroos, C. (1984): “A service quality model and its marketing implications”, *European Journal of Marketing*, 18 (4): 36–44.
- Howell, R.D.; Breivik, E.; Wilcox, J.B. (2007): “Is Formative Measurement Really Measurement? Reply to Bollen (2007) and Bagozzi (2007)”, *Psychological Methods*, 12 (2): 238–245.
- Janda, S.; Trocchia, P.J.; Gwinner, K.P. (2002): “Consumer perceptions of internet retail service quality”, *International Journal of Service Industry Management*, 13 (5): 412–431.
- Jarvis, C.B.; MacKenzie, S.B.; Podsakoff, P.M. (2003): “A critical review of construct indicators and measurement model misspecification in marketing and consumer research”, *Journal of Consumer Research*, 30:199–218.
- Johnson, R.A.; Wichern, D.W. (1998): “Applied multivariate statistical analysis”. Upper Saddle River, NJ: Prentice Hall.
- Kahn, J.H. (2006): Factor analysis in Counseling Psychology research, training and practice: Principles, advances and applications. *The Counseling Psychologist*, 34, 1–36.
- Kline, R.B. (2005): Principles and practice of structural equation modeling. New York: Guilford.
- Ladhari, R. (2008): “Alternative measures of service quality: a review”, *Managing Service Quality*, 18 (1): 65–86.
- Ladhari, R. (2009): “Service quality, emotional satisfaction, and behavioral intentions: A study in the hotel industry”, *Managing Service Quality*, 19 (3), 308–331.
- LeBlanc, G.; Nguyen, N. (1997): “Searching for excellence in business education: an exploratory study of customer impressions of service quality”, *International Journal of Educational Management*, 11 (2): 72–79.
- Lin, C.H.; Sher, P.J.; Shih, H.Y. (2005): “Past progress and future directions in conceptualizing customer perceived value”, *International Journal of Service Industry Management*, 16 (4), 318–336.
- Long, J.S. (1983): “Covariance structure models: an introduction to LISREL”, Sage University Paper Series on Quantitative Applications in the Social Sciences, N° 07-034, Newbury Park, CA: Sage.
- Markovic, S. (2006): “Students’ expectations and perceptions in Croatian tourism and hospitality higher education: SERVQUAL versus UNIQUAL”, *South East European Journal of Economics and Business*, 1 (1): 78–88.
- Martínez, R. (1995): “Psicometría: teoría de los test psicológicos y educativos”. Madrid: Síntesis Psicología S.A.
- Martínez-López, F.J.; Gázquez-Abad, J.C.; Sousa, C.M.P. (2012): “Structural Equation Modeling in Marketing and Business Research: Critical Issues and Practical Recommendations”, *European Journal of Marketing*, 47 (1): 1–33.

- Oh, H. (1999): "Service quality, customer satisfaction and customer value: a holistic perspective", *International Journal of Hospitality Management*, 18: 67–82.
- Parasuraman, A.; Zeithaml, V.A.; Berry, L.L. (1985): "A conceptual model of service quality and its implications for future research", *Journal of Marketing*, 49 (4): 41–50.
- Parasuraman, A.; Zeithaml, V.A.; Berry, L.L. (1988): "SERVQUAL: a multiple-item scale for measuring consumer perceptions of service quality", *Journal of Retailing*, 64 (1): 12–40.
- Parasuraman, A.; Zeithaml, V.A.; Berry, L.L. (1991): "Refinement and reassessment of the SERVQUAL scale", *Journal of Retailing*, 67 (4): 420–450.
- Podsakoff, N.; Shen, W.; Podsakoff, P.M. (2006): "The role of formative measurement models in strategic management research: review, critique and implications for future research". En Ketchen, D.J.; Bergh, D.D. (eds.): "*Research Methodology in Strategic Management*", Vol. 3. JAI Press, Amsterdam, pp. 197–252.
- Poza, C.; Fernández, J.A. (2011): "What Factors Explain the Multidimensional Poverty in Spain? An Approach by Means of Structural Equation Models", *Revista de Métodos Cuantitativos para la Economía y la Empresa*, 12: 81–110.
- Sharma, S. (1996): "Applied multivariate techniques". Nueva York: John Wiley & Sons, Inc.
- Snipes, R.L.; Thomson, N.F.; Oswald, S.L. (2006): "Gender bias in customer evaluations of service quality: an empirical investigation", *Journal of Services Marketing*, 20/4: 274–284.
- Spreng, R.A.; Mackoy, R.D. (1996): "An empirical examination of a model of perceived service quality and satisfaction", *Journal of Retailing*, 722: 201–214.
- Taylor, S.A.; Baker, T.L. (1994): "An assessment of the relationship between service quality and customer satisfaction in the formation of consumers' purchase intentions", *Journal of Retailing*, 70 (2): 163–178.
- Taylor S.A.; Nicholson, J.D.; Milan, J.; Valencia Martinez, R. (1997): "Assessing the roles of service quality and customer satisfaction in the formation of the purchase intentions of Mexican consumers", *Journal of Marketing: Theory and Practice*, 5 (1): 78–90.
- Teas, R.K. (1993): "Expectations, performance evaluation, and consumers' perceptions of quality", *Journal of Marketing*, 57 (4): 18–34.
- Ullman, J.D. (1996): "Structural equation modeling". En Tabachnick B.G. y Fidell L.S. (eds.): "Using multivariable statistics", New York: HarperCollins, 709–780.
- Uriel, E.; Aldás, J. (2005): "Análisis multivariante aplicado". Madrid: Thomson.
- Wang, Y.; Lo H.P. (2002): "Service quality, customer satisfaction and behavior intentions", *Info*, 4 (6): 50–60.
- Wilcox, J.B.; Howell, R.D.; Breivik, E. (2008): "Questions about formative measurement", *Journal of Business Research*, 61 (12): 1219–1228.
- Zeithaml, V.A.; Berry, L.L.; Parasuraman, A. (1996): "The behavioral consequences of service quality", *Journal of Marketing*, 60 (2): 31–46.

## ANEXOS

### A.1. BATERÍA PARA LA MEDICIÓN DE LA CALIDAD DE SERVICIO PERCIBIDA DE DOCENCIA

N°	Calidad de servicio percibida de docencia
<b>Tangibles</b>	
CS1	La modernidad y condiciones técnicas del equipo del aula son adecuadas para el desarrollo de la clase ( <i>data show</i> , computadora y accesorios, retroproyector, video, TV, etc.)
CS3	Las condiciones y elementos del aula son adecuados y funcionales para la docencia (pupitres, pizarrón, cortinas, enchufes, marcadores, borrador, etc.)
<b>Seguridad</b>	
CS18	El docente fomenta y respeta la expresión y participación de los estudiantes (Seguridad)
CS20	El docente es consistentemente cortés y respetuoso conmigo (Seguridad)
CS21	El docente mantiene un trato y respeto igualitario (Seguridad)
CS23	El docente promueve la comunicación docente-alumno (Seguridad)
CS24	El docente me da atención individualizada y personalizada (Empatía)
CS25	El docente es muy paciente y tolerante para explicar la materia (Empatía)
<b>Resultado</b>	
CS35	El grado de destrezas adquiridas (cognitivas, actitudinales, volitivas y conductuales) en la clase fue alto
CS36	El docente impartió la materia con un grado de complejidad adecuado
CS38	Los objetivos, la cobertura y profundidad de los temas han sido alcanzadas
CS39	Adquirí competencias profesionales (conceptuales, técnicas, humanas y competitivas) que puedo aplicar para resolver problemas y tomar mejores decisiones
CS40	La clase estuvo orientada al éxito, ya que se incrementaron mis capacidades de liderazgo y emprendedoras

### A.2. BATERÍAS PARA LA MEDICIÓN DE CONSTRUCTOS RELACIONADOS A LA CALIDAD DE SERVICIO DE DOCENCIA

N°	Calidad de servicio global de docencia
CSG1	El docente otorgó un excelente servicio de docencia global
CSG2	El docente otorgó un servicio de docencia de muy alta calidad
CSG3	El docente otorgó un alto estándar de servicio de docencia
CSG4	El docente otorgó un servicio de docencia superior en todos los sentidos
N°	Valor percibido del servicio de docencia
VAL1	La enseñanza que recibí del docente ha sido muy valiosa y útil
VAL2	Lo que recibí en esta clase vale más de lo que pagué
VAL3	Valió la pena dedicar mi tiempo a esta clase, por la enseñanza que recibí
VAL4	Valió la pena gastar mi energía y esfuerzo en esta clase, porque recibí una enseñanza de calidad
N°	Satisfacción global del servicio de docencia
SAT1	Estoy satisfecho de haber pasado clases con el docente
SAT2	Mi elección de pasar clases con el docente fue sabia
SAT3	Creo que hice lo correcto al cursar esta materia con el docente
SAT4	Creo que mi experiencia con este docente ha sido agradable
N°	Intenciones de comportamiento de postdocencia
IC1	Asistiría a otro curso (materia) impartido por el mismo docente
IC2	Recomendaría y animaría a mis compañeros a pasar clases con el docente
IC3	Diré cosas positivas acerca del docente a mis compañeros
IC4	Si fuera una autoridad académica, premiaría al docente por su desempeño
IC6	Permanecería en esta universidad hasta concluir mis estudios, si todos los docentes otorgaran el servicio de calidad que recibí en la clase