

MEDICIÓN DE LA CALIDAD DEL SERVICIO. DESARROLLO DE ESCALAS FRENTE A MEDIDAS DE UN SOLO ÍTEM. ESTUDIO COMPARATIVO.

Jose Antonio Martínez García,
Laura Martínez Caro

ABSTRACT

La medición de la calidad percibida es un tema recurrente en la literatura de marketing de servicios. La forma de obtener las evaluaciones del consumidor sobre el grado de excelencia del servicio preocupa tanto a académicos como a directivos. En esta investigación se han comparado las tres formas de medición comúnmente utilizadas en la literatura (escala multidimensional, escala multiatributo y escala de actitud global) frente a una medida de un solo ítem, con el fin de comprobar si existen diferencias sustantivas que apoyen la elaboración de cuestionarios menos costosos. El análisis se ha realizado en base a tres métodos de comparación: valores medios, comportamiento entre grupos de un factor, y correlación con una variable criterio. Los resultados muestran como la medida de un ítem se comporta de forma prácticamente idéntica que la escala multidimensional y la de actitud global en relación a los intervalos de confianza de los parámetros y al tamaño del efecto.

1. INTRODUCCIÓN

La calidad del servicio es considerada como uno de los determinantes clave del buen desempeño empresarial. Desde la década de los ochenta se han producido grandes cambios en el entorno de los negocios, siendo considerada la calidad como una de las prioridades para la dirección de cara a aumentar la competitividad y el crecimiento (Camisón, 1996; Devlin et al., 1994; Sureshghandar et al., 2002). Diversos estudios han mostrado amplias evidencias de que gestionar la calidad tiene gran importancia en la capacidad competitiva (Corbett, 1994, Kim 1995). Hoy en día, la calidad del servicio es reconocida como una de las cuestiones más importantes en el campo de la dirección de servicios y marketing, llegando la palabra “calidad” a ser parte del vocabulario diario de la dirección (Grönroos, 1990).

El crecimiento en la implantación de sistemas de aseguramiento de la calidad, como las normas ISO 9000 o los modelos de autoevaluación, es una muestra de la preocupación de las compañías por la mejora de los procesos y la calidad en la gestión. La versión vigente de las normas ISO 9000, la del 2000, pone mayor énfasis en la mejora continua, incidiendo en la necesidad de que el consumidor valore el desempeño del servicio. Asimismo los modelos de autoevaluación (EFQM¹, CAF², MC-OCSP³) enfocan parte de su análisis en la evaluación de los resultados sobre los clientes/ciudadanos y la sociedad en general. Esa evaluación conviene que se realice de manera sistemática, sirviendo como un instrumento que proporcione una información periódica del grado de excelencia del servicio, con el fin de obtener una retroalimentación constante acerca del nivel de desempeño organizacional. La necesidad de contar con instrumentos de medida adecuados se muestra, por tanto, fundamental para que las organizaciones puedan valorar la percepción de calidad del servicio desde el punto de vista del consumidor.

Sin embargo, las organizaciones se encuentran con diferentes disyuntivas a la hora de implementar esas mediciones. En primer lugar, decidir si pretenden conseguir valoraciones globales sobre

¹ European Foundation for Quality Management

² Common Assessment Framework

³ Modelo de Ciudadanía del Observatorio de Calidad de los Servicios Públicos

el desempeño o además obtener juicios sobre atributos o factores específicos del servicio que les proporcionen una información más detallada. En segundo término, escoger el instrumento de medición adecuado a esos objetivos de entre la variedad que la literatura propone. Y por último, diseñar el cuestionario definitivo y establecer el sistema de recogida de datos, considerando las dificultades en la obtención de la información y las limitaciones presupuestarias.

Dado que los costes de obtención y análisis de la información se incrementan con la longitud del cuestionario, y teniendo en cuenta que esos costes deben optimizarse, ya que las encuestas no deben ser una acción aislada sino sistemática, el objetivo de esta investigación es comprobar si existen diferencias en la evaluación de la calidad entre distintas escalas y una medida de un sólo ítem. De este modo, se pretende analizar si cuestionarios menos extensos podrían proporcionar una información igualmente válida a un menor coste. Así en tres contextos de estudio (servicio bancario, servicio deportivo y servicio de transporte urgente y mensajería), se compara la medición de la calidad con una escala diferente en cada sector (global, multiatributo y multidimensional) frente a una medida de un sólo ítem. Los resultados de los análisis comparativos y las implicaciones para la gestión son discutidos finalmente.

2. LA MEDICIÓN DE LA CALIDAD PERCIBIDA

La calidad percibida es un controvertido concepto en la literatura de marketing de servicios, y se identifica con un juicio personal y subjetivo que el consumidor emite sobre la excelencia o superioridad de un servicio o compañía (Bitner y Hubbert, 1994). Es una forma de actitud formada a partir de evaluaciones cognitivas y afectivas del individuo en su relación con la organización.

Desde la década de los ochenta, diversos modelos se han propuesto para medir la percepción de la calidad, confluyendo dos perspectivas claramente diferenciadas.

La primera de ellas plantea que los consumidores evalúan la calidad como el resultado de la divergencia entre las percepciones y las expectativas sobre el desempeño del servicio, tomando como base el paradigma de la disconfirmación de Oliver (1980). Siguiendo esta corriente, se encuentra el “Modelo Nórdico” desarrollado por Grönroos (1984), el “Modelo SERVQUAL” propuesto por Parasuraman et al. (1985, 1988) y posteriormente revisado (Parasuraman et al., 1994), el “Modelo de Tres Componentes de la Calidad del Servicio⁴” de Rust y Oliver (1994), o el “Modelo de Desempeño Evaluado⁵” planteado por Teas (1993, 1994).

La corriente de investigación que diverge de estos planteamientos defiende la superioridad de los modelos basados sólo en percepciones (Carman, 1990; Cronin y Taylor, 1992; McDougall y Levesque, 1994; Brady y Cronin, 2001). Esta perspectiva considera que incluir las expectativas puede ser ineficiente e innecesario, debido a que los individuos tienden a indicar consistentemente altos niveles de expectativas, siendo sus niveles de percepción raramente superior a esas expectativas (Babakus y Boller, 1992). Ello ha llevado al desarrollo de escalas alternativas al SERVQUAL, como el “SERVPERF” (Cronin y Taylor, 1992), la “Escala de Calidad del Servicio en un Entorno Minorista⁶” (Dabholkar et al., 1996) o el “Modelo Jerárquico Multidimensional” (Brady y Cronin, 2001).

Según el modelo definido por Brady y Cronin (2001) los consumidores forman sus percepciones sobre la calidad del servicio en base a una evaluación del desempeño en múltiples niveles, y al final combinan esas evaluaciones para llegar a la percepción global de la calidad del servicio. Los autores reconocen la complejidad del constructo y afirman que ninguna perspectiva es equivocada; cada una es incompleta sin la otra. A través de su investigación cualitativa y empírica, se muestra como la calidad del servicio conforma una estructura de tercer orden, donde la percepción de calidad es definida por claras y procesables dimensiones, y a su vez, éstas están constituidas por diversas subdimensiones. Brady y

⁴ En su nombre original: Tri- Component Model of Service Quality

⁵ Modelo “EP” (Evaluated Performance)

⁶ Escala “RSQS” (Retail Service Quality Scale)

Cronin (2001) sostiene que su propuesta de medición es la mejor forma de explicar la complejidad de las percepciones humanas teniendo en cuenta la conceptualización de calidad de servicio vigente en la literatura actual.

La calidad percibida es, de esta forma, una variable multidimensional, esto es, se manifiesta a través de una serie de constructos con un alto grado de correlación. Estos factores no son universales, tal y como algunos autores han sostenido (ej. Parasuraman et al., 1988) sino que son específicos del tipo de servicio evaluado (ej. Carman, 1990; Buttle, 1996). Para generar esos factores de calidad, se debe partir de estudios cualitativos en el caso de que no existan referencias en la literatura sobre ese sector específico y cultura concreta. De hecho, esa es la vía utilizada por diversos autores para generar mediciones en su contexto de estudio (Chumpitaz y Swaen, 2002).

De este modo, si se quieren obtener evaluaciones detalladas sobre diferentes atributos o factores de la calidad, se pueden construir modelos jerárquicos multidimensionales que proporcionen una visión estructural y multinivel de la calidad percibida. Estos modelos cuentan con un gran número de ítems, ya que los factores de calidad son variables latentes que se manifiestan a través de indicadores observables. La longitud del cuestionario depende del nivel de agregación del estudio: si se busca el nivel más desagregado (más detallado), la evaluación se realizaría a partir de las subdimensiones, mientras que si se busca un nivel más agregado (menos detallado) la evaluación se haría a partir de las dimensiones. En el primer caso (ej. Brady y Cronin, 2001; Ko y Pastore, 2005) la estructura es más compleja que en el segundo (ej. Fullerton, 2005).

Asimismo, esas escalas *ad hoc* pueden ser simplificadas considerando los atributos o factores de calidad como variables observables, sin carácter latente. Este enfoque parte igualmente de investigaciones cualitativas en conexión con la literatura relevante, generándose una serie de atributos o características evaluables del servicio, las cuales, normalmente, son estudiadas en su estructura factorial (ej. Bloemer et al., 1998). Esta forma de medición es menos consistente teóricamente que la de los modelos latentes multidimensionales, buscándose más una exploración de la estructura de los datos que una confirmación de la validez de la escala.

La calidad percibida es también medida de forma global, sin hacer referencia a los atributos específicos del servicio, como una evaluación de una actitud general sobre la excelencia del desempeño de la empresa. Esta opción de medición se suele realizar en investigaciones donde se analizan diferentes conceptos en redes de relaciones causa-efecto, a través de los modelos de ecuaciones estructurales. La medición global facilita la implementación de modelos causales, y permite una caracterización latente de la calidad percibida, pudiéndose estudiar su fiabilidad y validez. Sin embargo, se pierde información acerca de los atributos específicos del servicio, por lo que desde un punto de vista gerencial puede resultar menos atractivo. Numerosas investigaciones han contado con esta forma de evaluar la calidad (ej. Cronin et al., 2000, Brady et al., 2002).

Por último, varios autores han medido la calidad de forma global pero utilizando un único ítem (ej. Babakus y Boller, 1992; Bolton y Drew, 1991). Esta opción adolece de cierta consistencia psicométrica, ya que la calidad, al igual que otras variables que miden desempeño, como la satisfacción o la imagen corporativa, es considerada como un concepto abstracto, no directamente observable, y que conviene ser aproximado a través varios indicadores. Estos fenómenos abstractos son medidos con un cierto grado de fiabilidad, algo que no puede ser analizado con una única medida, por lo que ha de asumirse que el concepto está medido sin error⁷.

⁷ La utilización de modelos de ecuaciones estructurales permite fijar un error en las variables latentes medidas con un solo ítem, siempre que el modelo de relaciones propuesto cumpla las condiciones de identificación. Sin embargo, si simplemente se pretende estimar el valor medio de la calidad percibida en la población, sin considerar ninguna variable adicional, entonces ha de asumirse que esa la calidad está medida sin error.

3. ESTUDIOS COMPARATIVOS

Existen diversas investigaciones que han cotejado las propiedades psicométricas de distintas escalas multidimensionales, como por ejemplo entre los modelos SERVQUAL y SERVPERF (Cronin y Taylor, 1992), o entre el SERVPERF y el EP (Bou y Camisón, 2000). En ambos casos, la superioridad del modelo que opera sólo con percepciones (SERVPERF) fue probada empíricamente.

Asimismo, diferentes estudios han comparado escalas multidimensionales con medidas multi-ítem globales de calidad. Cronin et al. (2000), por ejemplo, muestran como existe una alta correlación (0.72) entre ambas variables, no hallando diferencias significativas en los valores medios de la calidad entre las dos formas de medición.

Otros autores han utilizado una medida de calidad global con un solo ítem como variable criterio para la validación de sus escalas multidimensionales (ej. Sierra et al., 2003), obteniendo también un alto nivel de correlación (0.77).

Sin embargo no se ha encontrado ningún estudio que plantee si la medida de un solo ítem puede ser una alternativa válida a las otras formas de medición. Se trata de pasar del enfoque tradicional en el que esta medida sólo es tenida en cuenta como criterio para validar otras escalas, a comprobar si es que realmente esta forma de medición produce resultados significativamente distintos de los que resultan de las escalas que son más comúnmente utilizadas en la literatura de calidad del servicio.

4. METODOLOGÍA

4.1 Ámbito del estudio e instrumentos de medida

La percepción de la calidad del servicio se estudió en tres contextos diferentes: servicio de transporte urgente y mensajería, servicio deportivo y servicio bancario, a través de escalas Likert de 1 a 5 puntos.

1. Servicio de transporte urgente y mensajería:

El objetivo de la investigación fue evaluar la calidad de este servicio en la Región de Murcia. Para ello se recogieron datos de una muestra de conveniencia de 200 sujetos que usaron este servicio en los últimos doce meses, delimitada sobre un marco muestral de 27000 individuos⁸. Con el fin de reforzar la representatividad de la muestra diversas cuotas referidas a variables demográficas como sexo y edad fueron respetadas. Adicionalmente, la distribución de empresas evaluadas se ajustaba a los datos sobre cuota de mercado entre empresas con una certificación de calidad y empresas no certificadas⁹.

La percepción de la calidad fue evaluada a través de un modelo jerárquico multidimensional de tercer orden desarrollado a partir de la revisión de la literatura (Brady y Cronin, 2001; Dabholkar, 1996; Ko y Pastore, 2005) y de una fase cualitativa con entrevistas personales a directivos y consumidores de este servicio. De esta forma la calidad percibida es considerada una evaluación global del servicio que se manifiesta a través de cuatro dimensiones: interacción personal, diseño del servicio, entorno físico y resultados del servicio, que a vez son evaluadas a través de una serie de factores más desagregados o subdimensiones: conducta, experiencia, resolución de problemas; rango de servicios, horario de envíos; aspectos tangibles, información; puntualidad, balance del servicio. Esas subdimensiones, consideradas variables latentes, fueron valoradas a través de una batería de 36 ítems¹⁰.

⁸ Cifra aproximada estimada partir de los datos de cuota de mercado proporcionados por directivos tras la realización de entrevistas personales.

⁹ Como indican Wilkinson y APA Task Force (1999), las muestras de conveniencia pueden ser "mejoradas" en relación a la representatividad de los datos si la muestra cumple con ciertas características de la población en un determinado rango de variables. De este modo, se trató de minimizar el sesgo que un muestreo de conveniencia produce sobre la estimación de los intervalos de confianza de los parámetros poblacionales.

¹⁰ El desarrollo y validación de esta escala puede verse con detalle en Martínez y Martínez (en prensa).

2. Servicio deportivo:

La calidad percibida del servicio municipal de las tres de piscinas de verano del Ayuntamiento de Cartagena (Murcia) fue evaluada a partir de una muestra aleatoria de 303 usuarios sobre un marco muestral de 1100 individuos. Se utilizó una escala multiatributo desarrollada a partir de una serie de entrevistas a técnicos deportivos y usuarios de piscinas, en conjunción con la revisión de la literatura (Luna y Mundina, 1998). Finalmente, siete atributos fueron considerados para medir la calidad: vestuarios, calidad del agua, higiene de la instalación, entorno físico, profesorado, personal de la instalación, y disponibilidad de horarios docentes. Estas variables fueron medidas a través de un sólo ítem cada una.

3. Servicio bancario:

El propósito de la investigación fue la evaluación comparativa de la calidad percibida del servicio de las dos cajas de ahorro con más cuota de mercado en la ciudad de Cartagena: Cajamurcia y CAM (Caja de Ahorros del Mediterráneo). El universo de estudio comprendió individuos de nacionalidad española con un rango de edad entre 16 y 64 años, acotándose un marco muestral de 43.403 sujetos¹¹. Se realizó un muestreo de conveniencia a 113 personas¹². La calidad fue medida como una actitud global sobre la excelencia del servicio a través de 3 ítems tomados de diversas fuentes (Brady y Cronin, 2001; Cronin et al., 2000; Olsen, 2002).

Medida de 1 ítem:

En los tres contextos de investigación, el encuestado valoró asimismo la calidad del servicio recibido a través de un solo ítem, medido en una escala Likert de 1 a 5 puntos: “Este/a (empresa/institución/entidad financiera) ofrece un excelente servicio”

4.2 Evaluación de las medidas

1. Servicio de transporte urgente y mensajería:

Se realizó un análisis factorial confirmatorio con LISREL 8.50 (Jöreskog y Sörbom, 2001) de las nueve subdimensiones del modelo, usando la matriz de covarianzas y una estimación por máxima verosimilitud. Dado el elevado número de parámetros a estimar con respecto al tamaño de la muestra se utilizó el procedimiento de disgregación parcial (ver Bagozzi y Heatherton, 1994; Dabholkar et al., 1996). El modelo presentó un ajuste exacto (Hayduk, 1996); $S-B\chi^2$:119.60; gl : 99; p : 0.077; Intervalo de Confianza para el $RMSEA$ al 90%: 0.00 – 0.05, así como adecuados niveles de fiabilidad y validez de constructo¹³. En una segunda estimación se testó el modelo completo de tercer orden obteniéndose un adecuado ajuste aproximado (MacCallum et al., 1996); $S-B\chi^2$:163.37; gl : 123; p : 0.008; Intervalo de Confianza para el $RMSEA$ al 90%: 0.021 – 0.056; $DELTA2$ y CFI : 0.973.

¹¹ Según el Instituto Nacional de Estadística (Padrón Municipal 2005), la población de Cartagena en ese rango de edad asciende a los 120.565 individuos. Dado que entre las dos entidades cuentan con un 36% de la cuota de mercado (Instituto Cajamar, 2001), el marco muestral se delimitó asumiendo que ese 36% de población puede ser cliente de una de esas dos entidades.

¹² La muestra final contempló a 59 usuarios de la CAM y 54 de Cajamurcia, en consonancia con la proporción de cuota de mercado según el número de sucursales (www.cajamurcia.es; www.cam.es).

¹³ Los valores de fiabilidad compuesta para las variables latentes estaban comprendidos entre 0.74 y 0.96, y los de la varianza media extraída (AVE) entre 0.59 y 0.93 (ver Bagozzi y Yi, 1988; Fornell y Larcker, 1981). Además, todos los ítems cargaron sobre sus respectivos factores, los parámetros estimados fueron entre 10 y 20 veces más grandes que sus errores estándar (Anderson y Gerbing, 1988). Por último el intervalo de confianza de las correlaciones entre constructos no incluyó el 1 en ningún caso. (Anderson y Gerbing, 1988).

2. Servicio deportivo:

Se exploró la estructura de los atributos (Bloemer et al., 1998) a través de un análisis de componentes principales con rotación oblicua Promax, obteniéndose 3 factores interpretables teóricamente que resumían un 71,5% de la varianza total. El primer factor agrupó atributos referentes a los elementos tangibles de la instalación: vestuarios, calidad del agua, higiene de la instalación, entorno físico (mínima carga factorial: 0.578; alpha de Cronbach: 0.722). El segundo factor englobó aspectos relacionados con la interacción personal: profesorado, trabajadores de la instalación (mínima carga factorial: 0.822). Por último, el rango de horarios docentes quedó representado en un tercer factor (carga factorial: 0.952).

3. Servicio bancario:

Esta conceptualización defiende que los indicadores de calidad son reflejo de un factor subyacente no observable. Por ello, se realizó un análisis factorial con extracción por máxima verosimilitud donde se obtuvo un factor que resumía el 60.2% de la variabilidad común (mínima carga factorial: 0.673; alpha de Cronbach: 0.801).

4.3 Métodos de comparación

Para comparar las tres formas de medición con la escala de un solo ítem, se halló la puntuación media final de cada una de las escalas, y se realizó una evaluación en base a tres aspectos: diferencia de medias, comportamiento entre grupos de un factor, y correlación con respecto a una variable criterio. De este modo se obtiene no sólo una comparativa sobre los valores medios de cada escala, sino también su capacidad para discriminar entre categorías de un factor y para relacionarse con otras variables de interés, por lo que el análisis sobre el grado de similitud de los resultados entre escalas es más completo.

Esa comparación es evaluada en base a los intervalos de confianza y los tamaños de efecto obtenidos. La utilización de estos dos criterios es altamente recomendable en los estudios comparativos (Rosnow y Rosenthal, 1989; Cohen, 1990, 1994; Schmidt, 1996), ya que contienen toda la información de los test de significación estadística, además de indicar la magnitud de las diferencias encontradas. Esta forma de mostrar los resultados ofrece una visión más amplia que los simples análisis de significación.

5. RESULTADOS

5.1. Test de diferencia de medias:

En primer lugar se hallaron los intervalos de confianza sobre la media muestral corrigiendo la varianza de la media por el factor de finitud de la población con el fin de obtener estimaciones más ajustadas. Se realizó una prueba T de muestras pareadas para generar un intervalo de confianza para la diferencia de medias y se estimó la magnitud de esa diferencia a partir del indicador de varianza explicada por el tratamiento (eta al cuadrado parcial), en este caso el instrumento de medida de calidad. Ese valor fue corregido por la correlación entre las medidas siguiendo las indicaciones metodológicas de Olejnik y Algina (2003) con el fin de evitar la sobreestimación del tamaño del efecto. Una vez obtenida esa corrección se estimó el tamaño del efecto f (Cohen, 1977) y la potencia del contraste a través del programa GPower (Faul y Erdfelder, 1992). Los resultados se muestran en la Tabla 1.

Tabla 1. Test de diferencia de medias.

	Media (IC al 95%)	IC al 95% para la diferencia de medias	Tamaño del efecto η^{2b}	Tamaño del efecto f^a	Potencia
Servicio transporte urgente		(-0.100 , 0.058)	0,0001	0.012	0.053
Escala jerárquica multidimensional	3.321 (3.229 , 3.413)				
Medida de un solo ítem	3.300 (3.169 , 3.431)				
Servicio deportivo		(-0.169 , -0.011)	0,004	0.065	0.203
Escala multiatributo	4.273 (4.213 , 4.333)				
Medida de un solo ítem	4.363 (4.290 , 4.436)				
Servicio bancario		(0.097 , 0.257)	0,013	0.113	0.224
Escala actitud global	2.831 (2.694 , 2.968)				
Medida de un solo ítem	3.008 (2.856 , 3.160)				

^a Convenciones de Cohen (1988) sobre el tamaño de efecto f : pequeño (0.1), mediano (0.25), grande (0.50).

^b Eta al cuadrado parcial corregida.

Los resultados muestran diferencias significativas en el servicio deportivo y servicio bancario (el intervalo de confianza de la diferencia de medias no contiene el cero), aunque los intervalos de confianza de los valores medios indican muy poca distinción sustantiva. Se obtienen tamaños de efecto muy pequeños para el servicio de transporte urgente y mensajería y el servicio deportivo, mientras que para el servicio bancario el efecto es mayor, aunque sigue siendo de pequeña entidad (0.113). De esta forma, la estimación de los valores medios de la calidad medida a través de un ítem proporciona puntuaciones prácticamente idénticas que las escalas de referencia.

5.2. Comportamiento entre grupos de un factor:

Se estudió el comportamiento de las escalas frente a un factor que a priori podría influir sobre los juicios de calidad. Para el sector de transporte urgente y mensajería, el factor escogido fue la certificación con la norma ISO 9000. De esta forma se dividió la muestra en dos grupos en función de la valoración de una empresa con la certificación o sin ella. En cuanto al servicio deportivo, las dos piscinas del centro urbano se agruparon en oposición a la tercera piscina situada en una pedanía de la ciudad, cuyos recursos son mucho más limitados. Con respecto al servicio bancario, se dividió a los encuestados en usuarios de Cajamurcia y CAM.

Se generaron los intervalos de confianza para las medias de cada grupo, estudiando su significación estadística a través de pruebas T de diferencia de medias de muestras independientes. La magnitud de la diferencia se estimó con la d de Cohen (Cohen, 1977) generándose asimismo intervalos de confianza para el tamaño de efecto según las indicaciones de Hedges y Olkin (1985). La potencia de los diferentes test se obtuvo con GPower.

Tabla 2. Comportamiento entre grupos de un factor

	Grupo 1 Media (IC al 95%)	Grupo 2 Media (IC al 95%)	IC al 95% para la diferencia de medias	Tamaño del efecto (<i>d</i> de Cohen) ^a	IC al 95% para el tamaño del efecto	Potencia
Servicio de transporte	Empresas certificadas	Empresas no certificadas				
Escala jerárquica multidimensional	3.569 (3.442 , 3.698)	3.148 (3.028 , 3.268)	(0.243 , 0.600)	0.668	(0.378 , 0.957)	0.996
Medida de un solo ítem	3.634 (3.446 , 3.814)	3.067 (2.896 , 3.238)	(0.312 , 0.821)	0.629	(0.341 , 0.918)	0.991
Servicio deportivo	Piscinas urbanas	Piscina pedanía				
Escala multiatributo	3.949 (3.819 , 4.079)	4.389 (4.326 , 4.452)	(-0.581 , -0.299)	0.746	(0.480 , 1,011)	0.999
Medida de un solo ítem	4.389 (4.255 , 4.523)	4.351 (4.264 , 4.438)	(-0.139 , 0.215)	0.051 ^b	(-0.208 , 0.309)	0.067
Servicio bancario	Cajamurcia	CAM				
Escala actitud global	2,833 (2.631 , 3.035)	2,831 (2.638 , 3.022)	(-0.273 , 0.279)	0.004	(-0.365, 0.373)	0.050
Medida de un solo ítem	2,944 (2.711 , 3.177)	3,068 (2.864 , 3.270)	(-0.429 , 0.182)	0.151 ^b	(-0.520 , 0.219)	0.125

^aConvenciones de Cohen (1988) sobre el tamaño de efecto *d*: pequeño (0.20), mediano (0.50), grande (0.80).

^b El tamaño del efecto se muestra en valor absoluto aunque la magnitud de la diferencia va en sentido contrario a la escala con la que se compara.

La Tabla 2 muestra los resultados de los diferentes análisis. En el sector de transporte urgente, la medida de un solo ítem se comporta de manera prácticamente idéntica a la escala jerárquica, tanto en la dirección del efecto como en el tamaño. Sin embargo, los resultados para los otros dos sectores son más ambiguos. En el caso del servicio deportivo existe una clara diferencia de comportamiento entre la escala multiatributo y la escala de un ítem. La primera discrimina de forma importante entre los factores (*d* de Cohen: 0.746), mientras la segunda no lo hace ni de forma significativa ni substantiva. En el servicio bancario, por su parte, las dos escalas muestran que no existen diferencias significativas en la percepción de la calidad entre las dos cajas de ahorro. Aunque la diferencia en el efecto tamaño de la escala de un ítem parece elevada con respecto a la escala global (0.151 frente a 0.004), no es así si se analizan los intervalos de confianza de las magnitudes de los efectos, obteniéndose valores mucho más parejos. No obstante, en este caso y en el de el servicio deportivo, existe la posibilidad de cometer un error tipo III, tal y como argumentan Leventhal y Huynh (1996); es decir, rechazar correctamente la hipótesis de que no existen diferencias entre ambos grupos pero inferir de forma incorrecta la dirección del efecto, algo que podría ser probable dado el bajo nivel de potencia estadística que presentan los contrastes que utilizan la medida de un solo ítem del servicio deportivo y servicio bancario, y la escala de actitud global de el servicio bancario. En cualquier caso, y teniendo en cuenta esta última consideración, la medida de un solo ítem produce resultados que en magnitud son sustantivamente similares que los de las dos escalas que operan con variables latentes: la jerárquica y la de actitud global.

5.3. Correlación con una variable criterio:

Por último se analizó la correlación de las escalas con respecto a una variable criterio. En los tres sectores esa variable fue un indicador de satisfacción del consumidor¹⁴. Para conseguir un análisis más completo de las comparaciones entre escalas se utilizó el software R2 (Steiger y Fouladi, 1992) en el que se pueden obtener los intervalos de confianza exactos de los coeficientes de correlación múltiple. Las correlaciones fueron comparadas utilizando el test Z de Steiger (Steiger, 1980) siguiendo las recomendaciones de Meng et al. (1992) sobre correlaciones que a su vez están correlacionadas. La Tabla 3 muestra los resultados de estos análisis.

Los resultados comparativos de las correlaciones con respecto a una variable criterio muestran como la medida de un solo ítem se comporta de nuevo de forma casi idéntica que la escala jerárquica, así como también con la escala de actitud global (los test Z de Steiger indican que no existen diferencias entre las correlaciones y los intervalos de confianza sobre R^2 son muy parejos). En el caso de la escala multiatributo, los coeficientes de correlación varían de forma significativa, aunque las diferencias en magnitud deben interpretarse con cautela debido al solapamiento de los intervalos de confianza del coeficiente de correlación múltiple (ver Schmidt, 1996).

Tabla 3. Correlación con una variable criterio

	Tamaño del efecto r (correlación de Pearson) ^a	Correlación ente las dos escalas de calidad	Z de Steiger	Coefficiente de correlación múltiple (R^2)	IC al 95% para R^2	Potencia
Servicio de transporte		0.808	-0.511			
Escala jerárquica multidimensional	0.815			0.665	(0.581 , 0.733)	1.000
Medida de un solo ítem	0.803			0.645	(0.557 , 0.717)	1.000
Servicio deportivo		0.550	- 2.895**			
Escala multiatributo	0.442			0.195	(0.118 , 0.279)	1.000
Medida de un solo ítem	0.574			0.329	(0.240 , 0.416)	1.000
Servicio bancario		0.797	0.000			
Escala actitud global	0.698			0.488	(0.346 , 0.609)	1.000
Medida de un solo ítem	0.698			0.488	(0.346 , 0.609)	1.000

^a Convenciones de Cohen (1988) sobre el tamaño de efecto para r : pequeño (0.10), mediano (0.30), grande (0.50).

** $p < 0.01$

¹⁴ La literatura sobre calidad del servicio muestra ampliamente la alta asociación entre calidad percibida y satisfacción del consumidor (Ej. McAlexander et al., 1994; Dabholkar, 1995)

6. DISCUSIÓN

La medida de un solo ítem ha mostrado un comportamiento prácticamente idéntico que la escala multidimensional y muy similar a la escala de actitud global en relación a los tres criterios de comparación propuestos en este estudio (valores medios, discriminación entre grupos y correlación con una variable criterio). Ambas escalas consideran a la calidad como una variable latente con error de medida, y aproximada por indicadores observables, aunque con diferente grado de complejidad. Estos resultados apoyarían la defensa de Hayduk (1996) y Hayduk y Glaser (2000) sobre la posibilidad de que un constructo latente pueda ser medido a través de un sólo indicador observable si ese ítem es lo suficientemente representativo del constructo latente subyacente. En este caso, el ítem escogido como única medida de calidad se comporta equivalentemente a dos escalas con buenas propiedades psicométricas.

Particularmente destacable es el resultado de la escala multidimensional, donde la calidad es desagregada en nueve subdimensiones medidas a través de 36 ítems. Dados los inconvenientes de una encuesta tan extensa desde el punto de vista del encuestado (cansancio, efecto aprendizaje, efecto halo), y desde la óptica del gestor (incremento de los costes), el hecho de obtener la misma información acerca del juicio global de calidad del servicio con una única pregunta puede ofrecer a las organizaciones nuevas alternativas a la hora de plantearse los estudios sobre calidad percibida del consumidor, siempre que estén dispuestos a renunciar a la información que detallada sobre factores de calidad que la escala multidimensional proporciona.

La escala multiatributo, sin embargo, muestra resultados menos claros. Aunque los valores medios de la calidad son prácticamente idénticos y las correlaciones con la variable criterio solapan sus intervalos de confianza, ambas mediciones se comportan de manera bastante diferente entre grupos de un factor. Este tipo de escalas están más asociadas a los análisis exploratorios sobre estructura de los atributos y carecen de la consistencia metodológica de los modelos jerárquicos multidimensionales en la construcción de las escalas. Es por ello que al no ser una herramienta psicométricamente tan aceptada, los resultados deben tomarse con más cautela. Además, dada la equivalencia de la medida de un ítem con las otras dos escalas, podría plantearse la cuestión de si es en este caso la escala multiatributo la que no proporciona una adecuada evaluación global de la calidad percibida.

Teóricamente, sin embargo, si la evaluación global del consumidor sobre la excelencia del servicio realizada sobre múltiples indicadores (escala de actitud global) coincide con la evaluación en un único ítem, el mismo razonamiento equivaldría para la utilización de múltiples dimensiones latentes (escala multidimensional), las cuales deberían ser también coincidentes con una medición de ellas a través de un solo ítem. Futuras investigaciones podrían realizar esa comparativa, de forma que se consiguieran simplificar las escalas jerárquicas multidimensionales sin perder información sobre los aspectos puntuales del servicio. Esa debiera ser la conceptualización de las escalas multiatributo, es decir, partir del enfoque riguroso de la modelización jerárquica para identificar las subdimensiones, y que esas subdimensiones sean en realidad los atributos a valorar.

Esta investigación ha mostrado que el sesgo producido en los valores medios por la medida de un solo ítem es muy pequeño, pudiendo ser perfectamente asumido por el investigador. De este modo, dada la importancia de la evaluación sistemática de la calidad percibida para las organizaciones, tanto para aquellas que adoptan sistemas formales de gestión de la calidad (certificación, modelos de autoevaluación), como para aquellas inmersas en procesos de modernización de los sistemas de gestión (sobre todo las administraciones públicas), el hecho de poder realizar una valoración continua de su desempeño a través de un único ítem podría facilitar enormemente esa evaluación cuando los recursos son más limitados. Esa medida de calidad no estaría exenta de error, aunque ese error no sea analizable, sino que más bien debería ser asumido o fijado por el investigador (Hayduk, 1996). Un inconveniente añadido sería la obtención de intervalos de confianza más amplios para los valores medios estimados de la calidad percibida, ya que su valor medio proviene de un rango discreto de valores para la valoración de

cada individuo (escala Likert), frente al valor medio obtenido de escalas multi-ítem, construido en base a valores medios continuos para cada sujeto (la media de varios ítems con rango discreto produce valores continuos). Ello repercutiría en estimaciones menos precisas por el incremento de la varianza de la media.

Recomendamos, finalmente, que siempre que no exista esa limitación de recursos se apueste por las escalas multidimensionales como una herramienta de diagnóstico más detallada que permita una valoración de puntos fuertes y débiles sobre los factores que conforman la calidad percibida.

7. LIMITACIONES

Esta investigación reconoce varias limitaciones. En primer lugar, hubiera sido más adecuado realizar todas las comparaciones en el mismo servicio para evitar que los resultados pudieran verse afectados por el contexto de investigación, sobre todo en términos de tamaño del efecto (Fern y Monroe, 1996). Sin embargo los tres estudios tenían objetivos específicos más amplios que la evaluación de la calidad percibida, siendo los cuestionarios bastante extensos. En segundo término, los intervalos de confianza para los parámetros del servicio de transporte urgente y de las entidades bancarias podrían verse ligeramente sesgados por la falta de completa aleatoriedad en el muestreo. Una tercera limitación proviene de la utilización de escalas Likert, las cuales plantean el problema de que la distribución muestral de cada ítem en muchos casos no es normal, caracterizada por altos grados de asimetría y curtosis, por lo que cuando la puntuación de la variable de interés no se obtiene de forma sumativa (es decir, si sólo existe un indicador para esa variable), algunos análisis estadísticos paramétricos pueden verse afectados en sus asunciones. Por último, y tal y como recomienda encarecidamente Cohen (1994), la replicación del estudio en otras muestras y contextos sería deseable con el fin de generalizar los resultados.

AGRADECIMIENTOS

Los autores agradecen a la Concejalía de Deportes del Ayuntamiento de Cartagena la aprobación para poder difundir los resultados del estudio sobre calidad en el servicio de piscinas de verano.

REFERENCIAS

- Anderson, J. C., y Gerbing, D. W. (1988). "Structural Equation Modeling in Practice: A review and recommended two-step approach". *Psychological Bulletin*, 103(3), 411-423.
- Babakus, E., y Boller, G. W. (1992). "An Empirical Assessment of the SERVQUAL Scale". *Journal of Business Research*, 24, 253-268.
- Bagozzi, R. P., y Heatherton, T. F. (1994). "A General Approach to Representing Multifaceted Personality Constructs: Application to State Self-Esteem". *Structural Equation Modeling*, 1 (1), 35-67.
- Bagozzi, R. P., y Yi, Y. (1988). "On the Evaluation of Structural Equation Models". *Journal of Academy of Marketing Science*, 16, (1), 74-94
- Bitner, M. J., y Hubbert, A. R. (1994). "Encounter Satisfaction Versus Overall Satisfaction Versus Quality", en Rust, R. y Oliver, R. (eds.) *Service Quality: New Directions in Theory and Practice*.
- Bloemer J., De Ruyter K., y Peeters P. (1998). "Investigating Drivers of Bank Loyalty: The Complex Relationship between Image, Service Quality and Satisfaction". *International Journal of Bank Marketing*, 16, 7, 276-286.
- Bolton, R., y Drew, J. (1991). "A Longitudinal Analysis of the Impact of Service Changes on Customer Attitudes". *Journal of Marketing*, 55 (January), 1-9.
- Bou, J. C., y Camisón, C. (2000). "Validity and Reliability in Perceived Quality Measurement Models. An Empirical Investigation in Spanish Ceramic Companies". *International Journal of Quality & Reliability Management*, 17 (8), 899-918.

- Brady, M. K., Cronin, J. J., y Brand, R. R. (2002). "Performance-Only Measurement of Service Quality: A Replication and Extension". *Journal of Business Research*, 55, 17-31.
- Brady, M. K., y Cronin J. J. (2001). "Some New Thoughts on Conceptualizing Perceived Service Quality: A Hierarchical Approach". *Journal of Marketing*, 5, 34-49.
- Buttle, F. (1996). "SERVQUAL: Review, Critique, Research Agenda". *European Journal of Marketing*, 30 (1), 8-25.
- Carman J. (1990). "Consumer Perceptions of Service Quality: An Assessment of the SERVQUAL Dimensions". *Journal of Retailing*, 66, 33-35.
- Chumpitaz, R., y Swaen, V. (2002). "Service Quality and Brand Loyalty Relationships: Investigating the Mediating Effect of Customer Satisfaction". *XXXI Conference European Marketing Academy*, Portugal.
- Cohen, J. (1977). *Statistical Power Analysis for The Behavioral Sciences*. New York: Academic Press.
- Cohen, J. (1988). *Statistical Power Analysis for The Behavioural Sciences* (2nd edition). Hillsdale, NJ: Erlbaum
- Cohen, J. (1990). "Things I Have Learned (so far)". *American Psychologist*, 45 (2), 1304-1312.
- Cohen, J. (1994). "The Earth Is Round ($p < .05$)". *American Psychologist*, 49 (12), 997-1003.
- Camisón C. (1996). "La Calidad como Factor de Competitividad en Turismo: Análisis del Caso Español". *Información Comercial Española*, 749 (Enero), 99-122.
- Corbett, L. M. (1994). "The Manufacturing Paradox: A Challenge for New Zealand Manufacturers". *Graduate School of Business and Government Management*, Special Report, No. 8, Victoria University, Wellington.
- Cronin, J. J., Brady, M. K., y Hult, G. T. M. (2000). "Assessing the Effects of Quality, Value, and Customer Satisfaction on Consumer Behavioral Intentions in Service Environments". *Journal of Retailing*, 76 (2), 193-218.
- Cronin, J., y Taylor, S. (1992). "Measuring Service Quality: A Re-Examination and Extension". *Journal of Marketing*, 56 (July), 55-68.
- Dabholkar P. A., Thorpe D. I., y Rentz J. O. (1996). "A Measure of Service Quality for Retail Stores: Scale Development and Validation". *Journal of the Academy of Marketing Science*, 24 (1), 3-16.
- Dabholkar, P. A. (1995). "A Contingency Framework for Predicting Causality between Customer Satisfaction and Service Quality". *Advances in Consumer Research*, 22, 101-108.
- Devlin, S. J., y Dong, H.K. (1994). "Service Quality from the Customers' Perspective". *Marketing Research*, 6 (1), 4-10.
- Faul, F., y Erdfelder, E. (1992). *GPOWER: A Priori-, Post Hoc-, and Compromise Power Analysis for MS-DOS* (computer program). Bonn, Germany: Bonn University.
- Fern, E. F., y Monroe, K. B. (1996). "Effect-Size Estimates: Issues and Problems in Interpretation". *Journal of Consumer Research*, 23, 89-105.
- Fornell, C., y Larcker, D. F. (1981). "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error". *Journal of Marketing Research*, 27 (February), 39-50.
- Fullerton, G. (2005). "The Service Quality-Loyalty Relationship in Retail Services: Does Commitment Matter?". *Journal of Retailing and Consumer Services*, 12, 99-111.
- Grönroos, C. (1984). "A Service Quality Model and Its Marketing Implications". *European Journal of Marketing*, 18 (4), 36-44.
- Grönroos, C. (1990). *Service Management and Marketing. Managing in the Moments of Truth in Service Competition*. Massachusetts: Lexington Books.
- Hayduk, L. A. (1996). *LISREL Issues, Debates and Strategies*. Baltimore, MA: Johns Hopkins University Press.
- Hayduk, L. A., y Glaser, D. N. (2000). "Jiving the Four-Step, Waltzing Around Factor Analyse and Other Serious Fun". *Structural Equation Modeling*, 7(1), 1-35.
- Hedges, L. V., y Olkin, I. (1985). *Statistical Methods for Meta-Analysis*. Orlando, FL: Academic Press.
- Instituto Cajamar (2001). "Boletín Económico Financiero Cajamar". *Instituto Cajamar*, III, 6.
- Instituto Nacional de Estadística (2005). *Padrón Municipal de Habitantes*. Instituto Nacional de Estadística.
- Jöreskog, K. G., y Sörbom, D. (2001). *LISREL 8.50*. Chicago: Scientific Software International.
- Kim, J. S. (1995). "Strategic Discoveries Trough Manufacturing Lessons from the 1994 International Manufacturing Futures Survey". *Proceedings Annual Conference Strategic Management Society*.
- Ko, Y. J., y Pastore, D. L. (2005). "A Hierarchical Model of Service Quality in the Recreational Industry". *Sport Marketing Quarterly*, 14 (2).

- Leventhal, L. y Huynh, C. (1996). "Directional Decisions for Two-Tailed Tests: Power, Error Rates, and Sample Size". *Psychological Methods*, 1, 278-292.
- Luna, R., y Mundina J. (1998). "La Satisfacción del Consumidor en el Marketing del Deporte". *Revista de Psicología del Deporte*, 13, 147-155.
- MacCallum, R. C., Browne, M. W., y Sugawara, H. M. (1996). "Power Analysis and Determination of Sample Size for Covariance Structure Modeling". *Psychological Methods*, 1(2), 130-149.
- Martínez, L., y Martínez, J. A. (en prensa). "Measuring Perceived Service Quality in Urgent Transport Service". *Journal of Retailing and Consumer Services*.
- McAlexander, J. H., Kadenburg, D. O., y Koenig, H. F. (1994). "Service Quality Measurement". *Marketing Health Services*, 14 (3), 34-44.
- McDougall, G. H., y Levesque, T. J. (1994). "A Revised View of Service Quality Dimensions: An Empirical Investigation". *Journal of Professional Service Marketing*, 11(1), 189-209.
- Meng, X. I., Rosenthal, R., y Rubin, D. (1992). "Comparing Correlated Correlation Coefficients". *Psychological Bulletin*, 111, 172-175.
- Olejnik, S., y Algina, J. (2003). "Generalized eta and omega Squared Statistics: Measures of Effect Size for Some Common Research Designs". *Psychological Methods*, 8 (4), 434-447.
- Oliver, R. L. (1980). "A Cognitive Model of the Antecedents and Consequences of Satisfaction Decisions". *Journal of Marketing*, 17 (November), 460-469.
- Olsen, S. O. (2002). "Comparative Evaluation and the Relationship between Quality, Satisfaction, and Repurchase Loyalty". *Journal of the Academy of Marketing Science*, 30 (3), 240-249.
- Parasuraman, A., Zeithaml V., y Berry L. (1985). "A Conceptual Model of Service Quality and its Implications for Future Research". *Journal of Marketing*, 49, 35-48.
- Parasuraman, A., Zeithaml V., y Berry 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., y Berry L. (1994). "Reassessment of Expectations as a Comparison Standard in Measuring Service Quality: Implications for Further Research". *Journal of Marketing*, 58 (January), 111-124.
- Rosnow, R. L., y Rosenthal, R. (1989). "Statistical Procedures and the Justification of Knowledge in Psychological Science". *American Psychologist*, 44 (10), 1276-1284.
- Rust, R., y Oliver, R. (1994). "Service Quality: Insights and Managerial Implications from the Frontier". In R.T. Rust y R.L. Oliver (eds) *Service Quality: New Directions in Theory and Practice*. Thousand Oaks, Ca; Sage Publications.
- Schmidt, F. L. (1996). "Statistical Significance Testing and Cumulative Knowledge in Psychology: Implications for Training of Researchers". *Psychological Methods*, 1 (2), 115-129.
- Sierra, B., Falces, C., Ruiz, M. A., y Alier, E. (2003). "Estructura de la Calidad Percibida en Servicios de Hostelería con Clientelas Cautivas". *ESIC MARKET*, Septiembre-Diciembre, 21-44.
- Steiger, J. H. (1980). "Test for Comparing Elements of a Correlation Matrix". *Psychological Bulletin*, 87, 245-281.
- Steiger, J. H., y Fouladi, R. T. (1992). "R2: A Computer Program for Interval Estimation, Power Calculation, and Hypothesis Testing for the Squared Multiple Correlation". *Behavior Research Methods, Instruments, and Computers*, 4, 581-582.
- Sureshghandar G. S., Chandrasekharan R., y Anantharaman R. N. (2002). "The Relationship between Management's Perception of Total Quality Service and Customer Perceptions of Service Quality". *Total Quality Management*, 13 (1), 69-88.
- Teas, R. (1993). "Expectations, Performance Evaluation, and Consumer's Perceptions of Quality". *Journal of Marketing*, 57 (October), 18-34.
- Teas, R. (1994). "Expectations as a Comparison Standard in Measuring Service Quality: An Assessment of a Reassessment". *Journal of Marketing*, 58 (January), 132-139.
- Wilkinson, L., y APA Task Force on Statistical Inference. (1999). "Statistical Methods in Psychology Journals: Guidelines and Explanations". *American Psychologist*, 54, 594-604.