

La evaluación ciudadana de la Administración local. La escala multidimensional jerárquica como instrumento para medir la calidad de los servicios locales¹

Enrique Carreras Romero

Universidad San Pablo CEU. Facultad de Económicas y Empresariales
ecr@mcalidad.es

Benjamín González Rodríguez

Universidad de A Coruña. Facultad de Sociología
mino@udc.es



Recibido: 15-02-2011
Aceptado: 27-04-2011

Resumen

Estamos asistiendo, en los últimos años, a un debate académico sobre la conveniencia y el papel que debería tener la evaluación del ciudadano en los sistemas de medición del desempeño o rendimiento (*performance*) de la Administración. Después de considerar las reservas de algunos autores, nos decantamos por la literatura que acentúa el carácter estratégico de la opinión del ciudadano. El artículo explora la viabilidad de una nueva escala para evaluar la calidad, percibida por el ciudadano, de los servicios de la Administración local. No hemos utilizado la escala SERVQUAL, por considerarla menos adecuada por varias razones, entre las que destaca la «incompatibilidad fáctica» de su aplicación. En su lugar, proponemos la escala multidimensional jerárquica como alternativa con más posibilidades. A partir de los datos de una encuesta realizada en varios municipios del Levante español, se ha procedido a realizar la validación empírica de la nueva escala, a fin de comprobar su capacidad predictiva sobre la intención favorable del ciudadano hacia la Administración local.

Palabras clave: medición del desempeño; administración local; escala SERVQUAL; escala multidimensional jerárquica; calidad percibida; evaluación ciudadana; modelos de ecuaciones estructurales.

1. Este trabajo ha sido posible, en parte, gracias a sendos proyectos de I+D+i (del MEC y del MICINN: SEJ2004-01098 y CSO2008-03337), dirigidos ambos por Benjamín González Rodríguez, coautor de este artículo.

Abstract. *Citizens' assessment of the local administration: The hierarchical multidimensional scale as an instrument for measuring the quality of local services*

In recent years, an academic debate has been launched on the usefulness and role that citizens' assessments should play in measuring public administration performance. Following a review of the literature, we focus on the strategic nature of citizens' views. This article explores the feasibility of a new scale, known as the hierarchical multidimensional scale, for assessing the quality of local government as perceived by the citizens themselves. The hierarchical multidimensional scale proves to be a much better alternative than the SERVQUAL scale due to the factual incompatibility in applying the latter scale. Using data from a survey conducted in several municipalities in eastern Spain, we empirically validate the new scale, and test its predictive ability on the positive intentions of citizens towards the local administration.

Keywords: measurement of performance; local administration; SERVQUAL scale; hierarchical multidimensional scale; perceived quality; citizen evaluation; structural equation models.

Sumario

- | | |
|--|---|
| <ol style="list-style-type: none"> 1. El debate sobre la evaluación del desempeño administrativo 2. La calidad percibida de los servicios de la Administración local 3. Metodología y datos 4. Hipótesis relativas a la validación empírica de la escala | <ol style="list-style-type: none"> 5. Resultados: La validación empírica de la escala multidimensional jerárquica 6. Discusión <p>Referencias bibliográficas</p> <p>Anexo: Cuestionario</p> |
|--|---|

1. El debate sobre la evaluación del desempeño administrativo

En los últimos años, está creciendo considerablemente la literatura sobre la metodología más adecuada para medir la eficiencia de la Administración pública. En concreto, el debate se centra en determinar el rol que debería realizar la evaluación del ciudadano en los sistemas de medida del desempeño administrativo (Stipack, 1979; Brown y Coulter, 1983; Hoogland DeHoog et al., 1990; Kelly y Swindell, 2002; Bouckaer and Van de Wallace, 2003; Licari et al., 2005; Kelly, 2005; Van Ryzin, 2006; Van Ryzin y Immerwahr, 2004; Van Ryzin et al., 2008; Shingler et al., 2008; Dalehite, 2008; Morgeson III y Mithas, 2009; James, 2009; Bruning, 2010; Howard, 2010).

La utilización de conceptos psicológicos, como «satisfacción» o «confianza», y sus efectos en el comportamiento electoral no es un tema nuevo y ha sido objeto de investigación en la literatura de la Administración pública (Lyons et al., 1992; Melkers y Thomas, 1998; Heintzman y Marson, 2005; Kampen et al., 2006; James y John, 2007), pero lo que resulta más novedoso en los escritos recientes es la consideración de la evaluación del ciudadano

como medida válida de la eficiencia administrativa (Van Ryzin et al., 2008; Shingler et al., 2008; Dalehite, 2008; Morgeson III y Mithas, 2009; James, 2009; Bruning, 2010).

Un renovado interés académico por los sistemas de medida del rendimiento público podemos encontrarlo en el movimiento de la Nueva Administración Pública (NPA). Esta corriente imagina la acción administrativa articulada alrededor de tres ideas básicas: el compromiso con la rendición de cuentas a la ciudadanía (*accountability*), la orientación de la acción administrativa hacia la eficiencia y la sensibilidad hacia el ciudadano administrado (Kamensky, 1996, Poister y Streib, 1999; Kettl, 2000; Yan y Hsieh, 2007; Kelly, 2007; Bourgon, 2007; Kuhlmann, 2010).

Kamensky (1996: 252) recoge una cita que sintetiza los criterios que deben regir la nueva gestión de los asuntos del gobierno:

La misión de esta revisión consiste en reinventar los sistemas de gobierno, rediseñar las agencias y el programa para que se ajusten mejor a sus clientes y dinamizar al gobierno. La tarea de reinventar el sistema provocará o producirá un marco para el desarrollo y la provisión de programas y políticas de coste eficientes por parte del Gobierno federal. Este marco clarificará la responsabilidad de los gestores sobre el logro de los resultados y creará un enfoque centrado en una nítida identificación y servicio al cliente y ofrecerá a los gestores las herramientas y los incentivos necesarios para concentrarse en los resultados.

Podríamos decir que la Nueva Administración Pública concibe los centros públicos como unidades de gestión que deben ser eficientes, transparentes y centradas en el ciudadano (Bourgon, 2007). La gestión requiere necesariamente la implantación de un sistema de medida del desempeño administrativo, capaz de evidenciar el grado de cumplimiento de los objetivos, tanto en proceso, como en resultados. Aunque la mayoría de los autores reconocen hoy la necesidad de dotarse de un sistema de evaluación del desempeño administrativo, el acuerdo desaparece cuando se propone incorporar la evaluación subjetiva del ciudadano como medida del rendimiento organizacional. En efecto, desde los primeros intentos por obtener sistemas de medida para monitorizar la acción administrativa, algunos autores han visto con recelo la utilización de las encuestas como sensores del rendimiento administrativo (Stipack, 1979, Brown y Coulter, 1983; Parks, 1984; Hoogland DeHoog et al., 1990; Lyons et al., 1992; Poister y Streib, 1999; Bouckaert y Van de Walle, 2003; Roch y Poister, 2006; Howard, 2010; Micheli y Neely, 2010). Howard (2010: 80) critica la utilización ingenua que se hace de las encuestas al ciudadano canadiense «Citizen First» y que no tienen en cuenta la complejidad de las interacciones entre satisfacción y realidad del servicio. Más decididos se muestran Bouckaert y Van de Walle (2003: 336) cuando afirman: «El buen gobierno descansa sobre una evaluación objetiva [...] esta evaluación tiene que descansar o asentarse sobre una serie de criterios (por ejemplo, no corrupción, no favoritismo por parte del gobierno, niveles adecuados de participación, etc.). Todos estos criterios son (inevitablemente) normativos».

Otros investigadores, por el contrario, se muestran partidarios de la evaluación ciudadana como fuente válida para juzgar la calidad de los servicios públicos. En los últimos años han aparecido diversos modelos para evaluar el desempeño público desde el punto de vista del ciudadano, tales como el «ACSI Government» (www.theacsi.org), en Norteamérica; el programa «Citizen First», en Canadá (Bruning, 2010; Howard, 2010), y «The Customer Voice», en el Reino Unido (www.cabinetoffice.gov.uk/chartermark). En España, desde el Ministerio de la Presidencia, se recomienda el uso de la escala SERVQUAL para la evaluación ciudadana de los servicios públicos (AEVAL, 2009).

Antes de seguir adelante con nuestra línea argumental, conviene destacar un matiz. Los críticos actuales con las encuestas no niegan el valor de conceptos subjetivos como la satisfacción o la confianza y su capacidad para explicar el clima de opinión o la conducta de voto, lo que ponen en tela de juicio es su validez para representar el desempeño administrativo (Bouckaert y Van de Walle, 2003; Howard, 2010). Por tanto, la polémica se circunscribe a la aceptación de los datos subjetivos *soft*, como medida de la calidad de un servicio público. En este sentido, encontramos autores que abogan por la preeminencia del dato subjetivo como medida válida del rendimiento administrativo (Kamensky, 1996; Kelly y Swindell, 2002; Kelly, 2005; Licari et al., 2005; Van Ryzin, 2006, 2007; Roch y Poister, 2006; James y John, 2007; Van Ryzin et al., 2008; Shingler et al., 2008; James, 2009; Morgeson III y Mithas, 2009). El debate se suscita alrededor de cuatro cuestiones, como son la falta de información, el sesgo de obligatoriedad del servicio, la escasa relación entre los datos *hard*, objetivos, y los *soft* provenientes de la subjetividad del ciudadano y la carga sociológica del ciudadano, sus características demográficas y la afinidad ideológica con el gobierno local (Stipack, 1979; Brown y Coulter, 1983; Parks, 1984; Hoogland DeHoog et al., 1990; Roch y Poister, 2006; Brown, 2007).

La primera objeción tiene que ver con el carácter limitado de la información disponible. El ciudadano no dispone de información suficiente para evaluar la eficiencia administrativa por dos razones: sólo controla los resultados que le son accesibles y tiene escasa capacidad discriminativa sobre el servicio (Stipack, 1979: 48). Ambos argumentos nos parecen poco sólidos. Obviamente, el ciudadano sólo suele tener acceso a los resultados de la acción administrativa y sólo en raras ocasiones controla los procesos o los recursos disponibles (Kuhlmann, 2010). Sin embargo, lo único que el ciudadano para emitir un juicio de valor es precisamente el aspecto visible del servicio, es decir, la transacción con el funcionario y su resultado. Argumentar que el desconocimiento del proceso invalida el juicio es, cuando menos, arriesgado, porque estaríamos solicitando que fuera «comprensible» o «generoso» con los resultados obtenidos. En todo caso esta actitud sería contraria al espíritu de la NAP que pretende orientar la gestión al ciudadano. En el segundo argumento (escasa discriminación) Brian Stipack hace una lectura parcial de la teoría de la asimilación-contraste de Anderson. Es cierto que las personas tenemos un umbral perceptivo indiferenciado, pero lo que el autor olvida es que, fuera de esos umbrales, el individuo diferencia con nitidez el nivel de prestación

de un servicio (mecanismo del «contraste», según Anderson, 1973: 40-41). Las personas vemos y diferenciamos los aspectos visibles y por tanto estamos facultados para establecer nuestras evaluaciones. Más adelante tenemos la ocasión de comprobar empíricamente la capacidad de los ciudadanos para discriminar entre los distintos servicios locales.

La segunda objeción para rechazar las evaluaciones del ciudadano procede de la obligatoriedad de algunos servicios públicos², como, por ejemplo, el pago de impuestos o los controles policiales (Kelly, 2005; Brown, 2007). Como señala Brown (2007: 568), los servicios de libre elección tienden a despertar valoraciones más favorables que los coercitivos, lo que lleva a provocar un «sesgo asimétrico» en las valoraciones. Nosotros pensamos que esta objeción no invalida la evaluación porque no afecta a la sensibilidad de la medida, sino al techo empírico que es razonable alcanzar. El doctor Kano ha detectado la presencia de asimetrías en las valoraciones de los consumidores (Oliver, 1997). Hay elementos del desempeño que son de obligado cumplimiento, y si se cumplen no despiertan valoraciones positivas, pero su incumplimiento actúa restando satisfacción. Obviamente, el techo de estos elementos será diferente de los elementos «atractivos» que sólo actúan incrementando la satisfacción, pero cuya ausencia no se asocia a la insatisfacción. Aunque los techos son diferentes, la variación en el servicio siempre se traslada a la valoración, sea en el rango bajo o alto de la escala, pero siempre sensible a los cambios reales del servicio.

La tercera objeción apunta a las incongruencias detectadas entre las medidas de las agencias y las evaluaciones del ciudadano en relación con los servicios locales, más próximos al ciudadano. En servicios como la policía, la limpieza de calles o las zonas de recreo, algunos autores encontraron una falta de correlación importante (Stipack, 1979; Brown y Coulter, 1983; Parks, 1984; Howard, 2010): «Encontramos una menor correlación entre las medidas de satisfacción de los ciudadanos y el *performance* administrativo de lo esperado» (Kelly y Swindell, 2002: 618). Parks (1984: 123) encontró que la relación entre registros y percepciones sobre la policía disminuía abruptamente cuando la persona no había tenido un encuentro reciente. Evidentemente, ante la duda, las medidas *hard* siempre parecen más fiables que las *soft* provenientes de la subjetividad del ciudadano. En realidad, la tercera objeción engloba funcionalmente a las anteriores. Si la información es limitada o sesgada, obviamente deberá detectarse una falta de congruencia entre los datos objetivamente registrados y las percepciones de los ciudadanos. De forma inversa, si se encontrara congruencia entre ambos tipos de datos, las objeciones anteriores tendrían poco fundamento. Investigaciones recientes sobre las administraciones locales han puesto en entredicho el desacuerdo entre los datos de los sistemas de medida objetivos y las percepciones de la ciudadanía (Kelly y Swindell, 2002; Licari et al., 2005; Van Ryzin et al., 2008).

2. Para un análisis detallado de las diferencias y analogías entre los servicios prestados por proveedores públicos o privados, nos remitimos a un manuscrito en vías de publicación de los mismos autores: González y Carreras, «El paradigma intencional en la evaluación ciudadana de la Administración local: La utilidad versus el bienestar».

Janet Kelly y David Swindell (2002: 618) encontraron que los indicadores *hard* y *soft* sobre resultados correlacionaban mejor que los de proceso o recursos. Por otro lado, Shingler et al. (2008: 1102) se percataron de que los ítems *hard* y *soft* comparados no medían la misma realidad, los registros de las agencias medían tiempos, errores o volumen de recursos, mientras que los ciudadanos valoraban aspectos más difíciles de registrar, como la sensibilidad del empleado, la inmediatez en atender o el interés demostrado.

Michael Licari, William McLean y Tom Rice idearon una investigación para dilucidar el grado de consistencia entre las percepciones del ciudadano y la realidad registrada. Las áreas de comparación serían las mismas, el mantenimiento de calles y parques y los datos de percepción del ciudadano serían contrastados con los datos de observación objetiva, obtenidos por un equipo experto. Los resultados fueron concluyentes: «Las encuestas a los ciudadanos son, hasta cierto punto, indicadores válidos de la calidad de las calles y de los parques» (Licari et al., 2005: 366). Tres años después, Van Ryzin et al. (2008: 301) replicaron una investigación similar combinando observaciones de expertos y encuestas sobre la limpieza en Nueva York, con idénticos resultados: «Estos resultados demuestran que los juicios de los ciudadanos sobre el *performance* del gobierno pueden corresponderse estrechamente con resultados medidos de una manera más objetiva». La evidencia actual parece que juega a favor de la capacidad ciudadana para juzgar la calidad de la acción administrativa, al menos en los resultados de servicios visibles. Sin embargo, en ambos estudios, también detectaron que los rasgos demográficos del ciudadano alteraban la valoración del servicio, lo que nos lleva a plantear la cuarta objeción contra la validez del dato perceptivo.

En efecto, esta cuarta objeción es importante, porque podría estar planteando una petición de principio. Si la ideología o la raza tuvieran un fuerte impacto, la valoración del servicio por parte del ciudadano no estaría reflejando la calidad percibida, sino la propia posición ideológica o la posición de raza. En ese caso, la evaluación ciudadana no debería tomarse como referencia para la gestión del servicio público. Las objeciones anteriores de información limitada, sesgo de obligatoriedad o baja correlación entre registros y encuestas no conculcan el principio de la gestión administrativa orientada al ciudadano, sin embargo, la cuarta objeción pone en suspenso dicho supuesto.

Por razones de espacio, hemos tenido que posponer la prueba empírica que invalida la cuarta objeción para otro artículo de próxima publicación en esta misma revista. En este sentido, daremos por supuesto que el ciudadano mantiene una capacidad discriminativa sobre el *performance* administrativo, por encima de su posición ideológica. Nuestro propósito en este artículo ha sido tratar de fundamentar la pertinencia empírica de una nueva escala para evaluar los servicios locales desde el punto de vista del ciudadano, mediante la explotación secundaria de los datos de una encuesta realizada en nueve ciudades del Levante español a finales de 2008. En concreto, hemos verificado las propiedades psicométricas de la escala multidimensional jerárquica de calidad percibida, así como los efectos, teóricamente esperados, en la formación de una

intención de conducta favorable hacia el gobierno local, expresada en el clima de opinión y la intención de voto.

La primera parte repasa las herramientas que se han utilizado para evaluar la calidad de los servicios, entre las que destacan la escala SERVQUAL (Parasuraman et al., 1988), la escala SERVPERF (Cronin y Taylor, 1992, 1994), la escala del *performance* evaluado (Teas, 1993, 1994) y la escala multidimensional jerárquica (Dabholkar et al., 1996). A pesar de la popularidad alcanzada por la escala SERVQUAL, incluso en el ámbito de la Administración (Van Ryzin, 2006; Roch y Poister, 2006; James y John, 2007; James, 2009), nos alejamos de ella para comprobar las posibilidades de una nueva escala, a nuestro juicio, más adecuada para los servicios locales, como es la escala multidimensional jerárquica.

2. La calidad percibida de los servicios de la Administración local

No cabe duda de que la ciudadanía mantiene y expresa ideas más o menos claras del grado de excelencia de la gestión local. Resulta habitual escuchar críticas o alabanzas a la intervención de la Policía Local, comentarios sobre la rapidez en la gestión de un trámite o la desgana con que la Administración ha atendido una queja. La opinión pública está llena de sugerencias sobre la mejora de los servicios, sobre la transparencia, conveniencia o justicia del reparto de prestaciones sociales o sobre la necesidad de subvencionar determinados proyectos. Los ciudadanos tienen ideas y las expresan con claridad meridiana.

Ya hemos comentado que algunos autores han recurrido a la escala SERVQUAL para evaluar la calidad de los servicios. Nosotros pensamos que resulta poco apropiada, dadas las dificultades de su aplicación a los servicios locales. Concretamente, hemos detectado cuatro inconvenientes:

- 1) Los problemas de validez de contenido (Nunnally y Berstein, 1994), que derivan de su incapacidad para abarcar la variedad de situaciones implicadas en los servicios locales.
- 2) La «incompatibilidad fáctica» en su aplicación.
- 3) Los mayores costes de campo.
- 4) La escasa validez predictiva (Cronbach y Meehl, 1955) sobre un constructo focal como la intención.

En primer lugar, la incapacidad de la escala SERVQUAL para integrar la diversidad de los servicios municipales. La mayoría de los ítems de la escala SERVQUAL han sido derivados de situaciones que implican un encuentro directo entre «empleado y cliente», como, por ejemplo, la «atención personalizada», la «exactitud en la respuesta», «no mostrarse ocupado», «confiar en el empleado», la «presencia física», etc. (Brady y Cronin, 2001; Ofir y Simonon, 2007). Estos ítems no resultan apropiados para servicios como mantenimiento de calles o jardines, que no requieren interacción directa con empleados.

La segunda cuestión, incompatibilidad fáctica, tiene que ver con la forma en que se construyen las puntuaciones SERVQUAL. Para obtener una medida

SERVQUAL de calidad, se requieren dos operaciones. En primer lugar, se solicita al entrevistado que indique su nivel de expectativa en algún aspecto del servicio y después se pide que evalúe el *performance* real del servicio en dicho aspecto (Zeithalm et al., 1990). La diferencia de ambas medidas da lugar a la medida SERVQUAL en cada individuo, por esa razón también se la conoce como «medidas sustractivas de calidad» (Oliver, 1997).

La implementación de estas medidas supone un dilema práctico de difícil solución. Bien la información se recopila secuencialmente en dos momentos separados por la vivencia del servicio: (1) cuestionario de expectativas, (2) experiencia de servicio, (3) cuestionario de percepción, o bien la información se obtiene a la vez: una vez vivido el servicio, en un mismo momento, se administra el cuestionario de expectativas e, inmediatamente después, el de la percepción. Si la información se obtiene en el mismo momento, después de la experiencia del servicio, estaríamos coleccionando «expectativas retrospectivas», es decir, incorporarían ya la actualización derivada de la experiencia ya vivida. Cuanto más actualizadas las expectativas, más tienden a asimilarse a la evaluación del *performance* y menos real resulta la medida de discrepancia objetiva. Enrique Carreras tuvo ocasión de comprobar, en la fase de piloto del índice ECSI (European Customer Satisfaction Index), que las expectativas retrospectivas saturaron en el mismo factor que la imagen generalizada del servicio (Carreras, 2000).

En el caso de obtenerse la información en dos administraciones distintas, mediadas por la experiencia de servicio, se dispondría de expectativas iniciales reales, pero, en este caso, podrían resultar no comparables con las valoraciones del *performance*, porque la propia experiencia habría modificado no sólo el nivel, sino también la estructura de referentes relevantes. La experiencia enriquece la valoración, lo cual genera nuevos matices en la percepción, nuevos atributos que no contemplaba el haz inicial de expectativas. James Carman aportó evidencia de este proceso de elaboración de nuevos referentes. En la investigación sobre los cuatro sectores, encontró que las estructuras factoriales pre y post no coincidían, se habían enriquecido con la vivencia del servicio. De los nueve factores analizados, sólo cuatro coincidieron. La reestructuración se había producido por la sensibilización y el aprendizaje del individuo. «After using the service, they were far more knowledgeable and their assessment of quality was both different and more clear» (Carman, 1990: 48).

El dilema de la «incompatibilidad fáctica» de la calidad quedaría planteado en los términos siguientes: Aunque podamos suponer que la desconfirmación objetiva de las expectativas por el *performance* percibido, sea un proceso real que regula los juicios de calidad, su verificación es imposible, porque si las expectativas y el *performance* se miden en la misma administración, las expectativas retrospectivas no se corresponden con las iniciales y la sustracción tendría un sesgo difícil de determinar. En el caso de cumplimentarse en la secuencia pre y post, los atributos relevantes de comparación habrían cambiado, con lo que se vuelve imposible la sustracción, porque las estructuras pre y post son diferentes.

El tercer argumento tiene que ver con los costes de campo. El cuestionario se multiplica por dos. Primero se pregunta por el servicio en términos de expectativa y luego se vuelve a preguntar por los mismos aspectos, pero ahora en términos de calificación del servicio. La fatiga siempre conlleva problemas de atención y fiabilidad en las respuestas (Carman, 1990).

La cuarta desventaja radica en su escaso valor predictivo. Estudios comparativos han demostrado la menor capacidad predictiva sobre la satisfacción que tiene la escala SERVQUAL frente a otros tipos de escalas (Tsé y Wilton, 1988; Cronin y Taylor, 1992, 1994; Brown et al., 1993; Teas, 1994), como la escala SERVPERF de Cronin y Taylor (1992), o la escala del *performance* evaluado de Teas (1993, 1994). Brown, Churchill y Peter (1993) han examinado los problemas métricos³ que están detrás de la pérdida de varianza explicada (R^2) en las escalas duales o sustractivas.

Las escalas de evaluación directa del servicio, tipo SERVPERF (Cronin y Taylor, 1992) son más ágiles, porque sólo necesitan la medida de valoración del servicio. Posiblemente, estén menos fundamentadas teóricamente (Parasuraman et al., 1994), pero han demostrado mayor capacidad predictiva en los modelos de satisfacción (Tsé y Wilton, 1988; Cronin y Taylor, 1992; Teas, 1993, 1994). Los propios autores del SERVQUAL reconocieron que la evaluación directa tenía mayor poder predictivo⁴. Otras investigaciones han corroborado este supuesto. Gilbert Churchill y Carol Surprenant determinaron que la valoración directa del *performance* en los aparatos de vídeo era el factor que más influía en la satisfacción, lo cual anulaba la influencia de las expectativas (Churchill y Surprenant, 1982), Fornell y colegas han encontrado trivial el impacto directo de las expectativas sobre la satisfacción (Fornell et al., 1996: 14). La calidad percibida y el valor percibido acumulan prácticamente todo el impacto sobre la satisfacción. En otros índices públicos de satisfacción, como el europeo, el sueco o el noruego, los resultados han corroborado la inoperancia del efecto directo de las expectativas (Johnson et al., 2001).

Richard Oliver ha propuesto una explicación teórica de las expectativas pasivas. En los casos estudiados, las expectativas activadas no tendrían impacto sobre la satisfacción del cliente, porque serían «expectativas pasivas» (Oliver, 1997). La prestación continua de un servicio hace que el consumidor las codifique permanentemente y las vaya adaptando al nivel del *performance*. A menos que se produzca un suceso inesperado, el consumidor no volverá a activarlas. En nuestro caso, muchos servicios locales, como parques y jardines, complejos deportivos, teatros, casas de cultura o la regulación del tráfico urbano, caen dentro de este tipo de servicios. Los ciudadanos, en su vida diaria, suelen

3. Para Brown et al. (1993), las escalas duales sustractivas plantean problemas de restricción de la varianza debidos fundamentalmente a las medidas de expectativas, además de una menor fiabilidad y una menor capacidad discriminativa que las escalas de desconfirmación subjetiva y SERVPERF. Estos problemas de medida estarían detrás de la menor capacidad predictiva sobre ítems de satisfacción e intención de recompra.
4. «La varianza explicada es el único criterio sobre el que la escala SERVPERF funciona mejor» (Parasuraman et al., 1994: 116).

presenciar mayor continuidad en la prestación del servicio que situaciones inesperadas. La cotidianeidad de las vivencias hace que las evaluaciones del servicio se establezcan y adquieran una naturaleza actitudinal (Johnson et al., 2001). La calidad percibida de los servicios locales no sería más que imágenes estabilizadas y, por esa razón, expresarían actitudes sociales ante el servicio.

En definitiva, la escala SERVQUAL no es específica de cada servicio, es más cara de implementar, y su «incompatibilidad fáctica» vuelve estéril su utilización, tiene poco poder predictivo sobre la intención y la vivencia cotidiana de los servicios y vuelve inoperante uno de sus componentes: las expectativas.

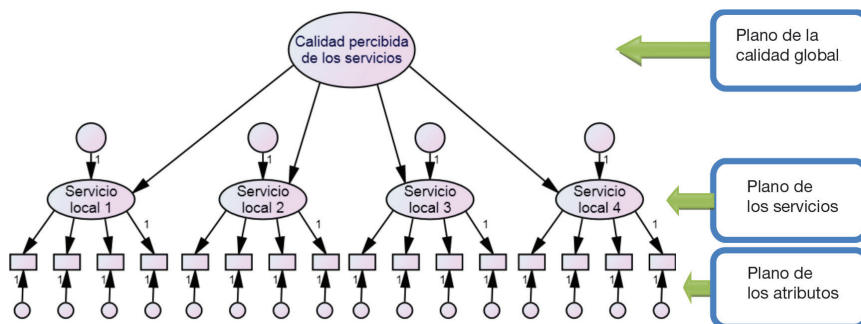
Creemos que una forma más adecuada de abordar la operacionalización de la calidad percibida de los servicios locales es la escala multidimensional jerárquica (Dabholkar et al., 1996; Brady y Cronin, 2001) mediante *evaluación directa* (Cronin y Taylor, 1992) y con *respuesta bipolar adjetival*, tal y como aconsejaron los propios autores del SERVQUAL (Parasuraman, 1994).

La escala multidimensional jerárquica (figura 1) se despliega en tres planos: el plano de los atributos o indicadores de calidad, el plano de los servicios locales y el plano de la calidad global percibida. Este tipo de escala responde a los requerimientos del servicio local y presenta varias ventajas. En primer lugar trata de forma independiente cada servicio, lo que permite que los atributos se adecuen a la especificidad de cada uno. Los atributos son medidos de forma directa, sin pasar por las expectativas y, además, la doble articulación permite comprobar el grado en que las evaluaciones concretas de servicios van generando una idea global de calidad. La investigación ha permitido comprobar la validez y utilidad de esta escala.

3. Metodología y datos

Se ha realizado un análisis secundario sobre los datos de una encuesta por muestreo probabilístico con entrevista telefónica, sobre calidad percibida de los servicios locales en nueve ciudades entre 25.000 y 100.000 habitantes, de

Figura 1. Esquema de la Escala Multidimensional Jerárquica



Valencia (España), dirigida por los profesores Juan Carlos Hortelano y Enrique Carreras, de la Universidad de Valencia y San Pablo CEU de Madrid, respectivamente. El trabajo de campo tuvo lugar entre el 23 de septiembre y el 20 de octubre de 2008. Los nueve municipios estaban gobernados por el mismo partido político. La presencia de varios municipios no invalida el modelo causal, porque la unidad de análisis es el individuo. Cada persona evalúa la calidad que percibe y desarrolla una intención.

Muestra: Se obtuvieron 570 respuestas completas, extraídas mediante muestreo sistemático a partir del listado de teléfonos fijos de los nueve municipios. El tipo de muestreo fue estratificado y proporcional por municipio, con tamaños intramunicipio que oscilaban entre cincuenta y ochenta casos. Dentro de cada municipio, se equilibraron las poblaciones por sexo y edad. El índice de penetración del teléfono fijo en los hogares principales para la Comunidad Valenciana en 2008 era del 75,7% (datos del Instituto Nacional de Estadística (INE): www.ine.es). El sesgo de cobertura (Groves, 2009) se ha controlado equilibrando la población entrevistada por sexo y edad, para que coincidiera con los datos oficiales de población en los nueve municipios (datos INE, 2008). La distribución muestral por variables demográficas fue la siguiente: varones de 18 a 45 años, 26,7%; varones mayores de 45 años, 23,2%; mujeres de 18 a 45 años, 24,6%; mujeres mayores de 45 años, 25,5%. Con el objeto de proteger la confidencialidad de la información generada, no se presentará ningún valor identificativo del municipio o de la persona, y las variables retenidas son las estrictamente necesarias para el desarrollo de la investigación, las de calidad percibida, intención hacia el gobierno local, sexo, edad e ideología⁵.

Selección de los servicios locales: aunque el estudio contemplaba once servicios municipales, se pactó con la empresa que cedería información sobre cuatro servicios municipales. Para obtener un abanico representativo, los servicios fueron seleccionados en función de las dimensiones que los caracterizan, como son el carácter obligatorio o voluntario de los servicios, la presencia o ausencia de interacción con empleados, la prestación en régimen de monopolio o de libre elección y el pago por servicios que no utilizan (Hyde, 1991; Kelly, 2005, Brown, 2007). Así, se pactó el Servicio de Policía Local, definida por el carácter obligatorio del contacto⁶; los servicios de uso voluntario sin interacción, como el de Parques y Jardines⁷; los que se prestan en régimen de monopolio, como los servicios de limpieza, y los trámites ante el Ayuntamiento que requieren una interacción tanto voluntaria, como en el caso de consultas o sugerencias,

5. Agradecemos a Medición de Calidad, SL el habernos facilitado los datos para realizar el análisis secundario que ha dado lugar al presente trabajo. La base de datos facilitada contiene únicamente la información necesaria para el desarrollo de la investigación, como son las variables de calidad percibida, y de intención de voto, y como datos de clasificación, el sexo, la edad y la ideología declarada. El resto de información ha sido excluida.
6. El 10,4% de la población entrevistada manifiesta haber tenido una disconformidad con la policía local en el último año, lo que supone una interacción obligada reciente.
7. El 35,5% de la población entrevistada dice no visitar un parque nunca o casi nunca, por tanto, paga impuestos por un servicio que no consume.

como compulsiva, cuando los trámites se relacionan con el pago de impuestos, las multas o los requerimientos administrativos⁸. Los servicios seleccionados para el estudio han sido:

- Los servicios de limpieza, gestionados empresarialmente.
- Los Parques y Jardines, servicio sin contacto directo con empleados, ni contratación.
- La Policía Local, acción pública proactiva y reactiva ante la demanda del ciudadano.
- La gestión administrativa del Ayuntamiento: consultas, sugerencias, quejas, trámites, subvenciones, ayudas, impuestos, multas, licencias.

Pensamos que los cuatro servicios garantizan cierta diversidad, cubren las especificidades propias de los servicios locales y son de alto reconocimiento social, es decir, son susceptibles de generar opinión, aun cuando el ciudadano no los haya experimentado directamente.

Cuestionario

Los ítems para evaluar la calidad percibida de los servicios son específicos para cada servicio y no responden a un esquema común de indicadores como en el cuestionario SERVQUAL. La identificación de los ítems de calidad corrió a cargo de la empresa depositaria de los datos. Realizó quince entrevistas en profundidad a ciudadanos de la ciudad de Valencia. Se exploraron aspectos relacionados con experiencias satisfactorias e insatisfactorias con cada servicio. Se retuvieron los ítems con más capacidad para aumentar o disminuir la satisfacción del ciudadano (Oliver, 1997). Un análisis factorial exploratorio redujo la relación de los 103 ítems iniciales identificados para los once servicios a treinta y tres ítems. Por otra parte, de los dieciséis indicadores que medían la calidad de los cuatro servicios contemplados en este estudio, sólo utilizamos los once que mejor se adaptaban a los requerimientos técnicos del análisis factorial confirmatorio. El cuestionario utilizado en este estudio está disponible en el anexo.

Análisis

Para explorar la viabilidad empírica de la escala multidimensional jerárquica aplicada a los servicios locales, se realizó un análisis factorial confirmatorio. Este método ha sido ampliamente utilizado en la literatura para verificar la capacidad psicométrica de una escala (Anderson y Gerbing, 1988; Cronin y Taylor, 1994; Spreng et al., 1996; Cronin et al., 2000; Dabholkar et al., 1996;

8. Del 43,8 % de las personas que tuvieron un encuentro con el Ayuntamiento para realizar una consulta, una sugerencia, solicitar ayuda o realizar una gestión, el 16,8% fue en relación con multas, impuestos o tasas.

Brady y Cronin, 2001; Sirdeshmukh et al., 2002; Harris y Goode, 2004; Lam et al., 2004; Ngobo, 2004; Cater y Zabkar, 2008).

4. Hipótesis relativas a la validación empírica de la escala

En concreto, la presencia empírica de un constructo multidimensional jerárquico en la calidad de servicios locales requiere la verificación de las siguientes hipótesis:

- H₁: Los constructos de los servicios locales son fiables y tienen suficiente validez extractada (validez convergente).
- H₂: La escala discrimina entre los servicios locales (validez discriminante).
- H₃: La calidad global percibida de los servicios es un constructo de segundo orden que se origina a partir de las evaluaciones de los servicios locales concretos.
- H₄: La calidad global percibida de los servicios tiene validez nomológica, mantiene una relación lineal positiva con la intención favorable al gobierno local (comentarios positivos e intención de voto).

Las dos primeras hipótesis someten a prueba la capacidad de la escala para representar evaluaciones de servicios locales concretos y diferenciados entre sí. La tercera hipótesis comprueba el grado en que la ciudadanía mantiene una idea global de calidad percibida, construida a partir de las evaluaciones de servicios locales concretos y diferentes. Por último, la cuarta hipótesis testa la validez nomológica de la calidad percibida o su capacidad para ser antecedente de la intención hacia el gobierno local. La propiedad de convergencia de los ítems que representan un servicio (H₁) se verifica mediante el cálculo de los índices de fiabilidad y validez extractada (Fornell y Larcker, 1981; Anderson y Gerbing, 1988). La capacidad de discriminación (H₂) se prueba mediante dos procedimientos consecutivos, la restricción de las correlaciones entre constructos a «1» y el cálculo de los intervalos de confianza de dichas correlaciones (Anderson y Gerbing, 1988). La hipótesis de la calidad global como constructo de segundo orden (H₃) ha sido comprobada contrastando la pérdida de ajuste de dicho modelo respecto del modelo de covarianzas saturado (Dabholkar et al., 1996; Brady y Cronin, 2001). Por último, la validez nomológica (H₄) o capacidad de la escala para explicar la intención hacia el gobierno local (Anderson y Gerbing, 1988) se ha planteado mediante el ajuste de un modelo causal (Bollen, 1989).

5. Resultados: La validación empírica de la escala multidimensional jerárquica

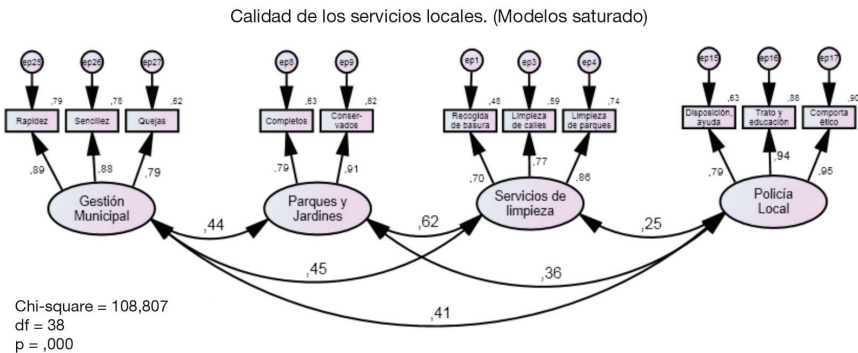
El análisis factorial confirmatorio se ha ajustado por el método de máxima verosimilitud (programa AMOS). La técnica de ajuste es adecuada por las siguientes razones:

- Los datos mantienen distribuciones individuales con coeficientes de asimetría que varían entre $-1,193$ y $-0,278$ y curtosis entre $-0,868$ y $+1,663$. Aunque el índice de normalidad multivariada resulta significativo (con una curtosis significativa con un ratio de coeficiente del $37,141$), el análisis de los casos *outliers* indica que están agrupados sin elementos con grandes desviaciones (Kline, 1998). Consideramos que el método de máxima verosimilitud es adecuado a los datos que manejamos
- En la estimación de los parámetros no se han detectado varianzas negativas y la cantidad de casos es suficiente para la estimación. Los 570 casos están por encima de los 180 casos requeridos (18 variables por 10) (Bagozi y Yi, 1988).
- No hay varianzas negativas en los errores estimados y todas ellas son significativas (Bollen, 1989).

Para verificar las propiedades del modelo de medida, la representación de las latentes por sus indicadores respectivos, se recomienda saturar las covarianzas entre las latentes, de forma que la falta de ajuste sería imputable a dicho modelo de medida (Bagozi y Yi, 1988; Anderson Gerbing, 1988). La figura 2 contiene los resultados del modelo saturado de los cuatro servicios municipales seleccionados, con sus indicadores respectivos.

Bajo la hipótesis nula de reproducción de la matriz poblacional de covarianzas observadas, « Σ » a partir de la matriz poblacional de las covarianzas implicadas por el modelo « $\Sigma(\theta)$ » $\rightarrow (H_0: \Sigma(\theta) = \Sigma)$. El ajuste del modelo presenta una discrepancia total, $\chi^2 = 108,807$, con 38 grados de libertad. El *p-valor* parece rechazar la hipótesis nula con una probabilidad de error prácticamente inexistente, del 0,000, con lo que parecería concluir que los datos no se ajustan al modelo propuesto.

Figura 2. Resultados de la estimación de los parámetros



Fuente: Elaboración propia.

Tabla 1. Índices de bondad del ajuste

GFI	CFI	SRMR	RMSEA	Pclose _(0,06)
0,967	0,982	0,033	0,057	0,626

Fuente: elaboración propia.

Sabemos que el contraste basado en la χ^2 presenta cierta ambigüedad debido al tamaño de muestra⁹, lo que lleva a rechazar ajustes triviales o bien a aceptar desajustes importantes. Por esa razón, se aconseja valorar el ajuste con otros índices (Anderson y Gerbing, 1998; Bagozzi y Yi, 1988; Bollen, 1989; Hu y Bentler, 1999). Para valorar el ajuste, seguimos la estrategia de doble índice¹⁰ de Li-Tze Hu y Peter Bentler (1999). Para modelos estimados por máxima verosimilitud, los índices seleccionados y sus puntos de corte son: GFI (> 0,95) y SRMR (< 0,11) y el CFI (> 0,95) y RMSEA¹¹ (< 0,06). Según los resultados, el ajuste del modelo saturado es excelente (tabla 1):

El *Pclose* ha sido calculado sobre la hipótesis nula de que el RMSEA no sea mayor de 0,06¹². Una vez validado el ajuste del modelo, pasamos a comprobar las hipótesis de convergencia y discriminación de los constructos.

Validez convergente de los constructos

La validez convergente se determina con el índice de fiabilidad y el promedio de varianza extractada AVE. Los valores aceptables son, para la fiabilidad, > 0,6 y, para el AVE, > 0,5 (Fornell y Larcker, 1981; Bagozzi y Yi, 1988). En la tabla 2, se ve que todos los valores superan los niveles mínimos requeridos. Los índices de fiabilidad y promedio de varianza extractada (AVE) superan los criterios de aceptación. Los resultados corroboran la validez convergente de los indicadores que representan cada uno de los cuatro servicios locales (H_1).

9. Recordemos que no es el valor total de discrepancia F lo que se comporta como una función de probabilidad χ^2 , sino la discrepancia multiplicada por el tamaño de muestra ($n \cdot F$). Por esa razón, el tamaño posibilita el contraste, a la vez que lo condiciona (Bollen, 1989).
10. Los autores encontraron dos grupos de índices altamente correlacionados entre sí. El primero (NFI, BL86, GFI, AGFI, CAK, SRMR y CK) se mostraba más sensible con el modelo de impacto mal especificado, y el segundo (TLI, BL89, RNI, CFI, Mc, RMSEA), más sensible con las cargas factoriales mal especificadas. Si se eligieran representantes de ambos grupos, se conseguiría una valoración razonable del ajuste (Hu y Bentler, 1999).
11. Adjuntamos el contraste de Browne y Cudeck (1993) que prueba la hipótesis nula de que el RMSEA sea 0,05 o menos (H_0 : RMSEA \leq 0,05). Dado que el P-valor es 0,162, no rechaza hipótesis nula.
12. La hipótesis de contraste sería: H_0 : RMSEA \leq 0,06. El *Pclose* da la probabilidad de rechazar la hipótesis nula siendo verdadera. $Pclose = 1 - \phi$, siendo ϕ la probabilidad de la función chi-square no central, con valor crítico = χ^2 ; parámetro no central = $0,06^2 \cdot n \cdot g.l.$; con los grados de libertad del modelo.

Tabla 2. Índices de validez convergente por servicio local

Servicios	Fiabilidad	"AVE" ^(*)
Policía Local	0,92	0,80
Servicios de limpieza	0,82	0,60
Parques y Jardines	0,84	0,73
Gestión municipal	0,89	0,73
Satisfacción con la gestión	0,93	0,93

(*) Acrónimo en inglés: promedio de varianza de los indicadores extractada por la latente.
Fuente: Elaboración propia.

Validez discriminante de los constructos

La validez discriminante o capacidad del modelo para apuntar constructos diferentes se comprueba mediante los dos procedimientos ya comentados: la restricción de la correlación entre pares de latentes y el cálculo de los intervalos de confianza de las correlaciones entre dichos pares (Anderson y Gerbing, 1988). La correlación entre cada par de latentes está restringida a la unidad¹³. Si, al forzar la correlación a uno, la pérdida de ajuste respecto del modelo saturado es significativa, concluimos que los datos no aceptan que las latentes sean iguales. Como los datos verifican una pérdida importante de ajuste entre todos los pares de servicios locales, se rechaza que las latentes sean las mismas (tabla 3).

El segundo procedimiento proporciona una idea del grado de discriminación entre las latentes. Consiste en calcular el intervalo de confianza de las correlaciones estimadas. Cuanto más se aleje el límite superior del intervalo de la unidad, mayor capacidad de discriminación habrá entre el par bajo prueba. La tabla 4 muestra los resultados de los intervalos de confianza, al nivel del 95%, de las estimaciones de las correlaciones.

Los intervalos muestran valores del límite superior alejados de la unidad, lo cual indica una discriminación aceptable entre los servicios. Ambos resultados corroboran la H_2 : los servicios locales despiertan evaluaciones diferentes entre la ciudadanía.

La tercera hipótesis prueba la presencia de una idea global de calidad de los servicios locales. Técnicamente, consiste en verificar la pérdida de ajuste entre un modelo de segundo orden y el modelo de covarianzas saturado (Dabholkar et al., 1996; Brady y Cronin, 2001). Examinando los resultados de los ajustes de ambos modelos (figura 3), encontramos diferencias significativas, lo que indicaría una pérdida de ajuste al introducir el constructo de segundo orden. Sin embargo, debemos valorar la pérdida de ajuste, ya que el estadístico de contraste χ^2 es sensible al tamaño de la muestra. Fang Chen propone utilizar

13. Para restringir la correlación entre dos latentes a la unidad en el programa AMOS, se deben fijar tres parámetros, la covarianza del par de latentes, el producto de sus desviaciones típicas (correlación = 1) y las varianzas de dichas latentes a sus valores estimados. De este modo, el modelo anidado tendrá tres grados más de libertad y el modelo ajustado dará como correlación la unidad.

Tabla 3. Validez discriminante. Pruebas de restricción de la correlación a la unidad

Modelo no restringido						
Chi-square =		108,807				
Grados de libertad (g.l.)		38				
Modelos restringidos						
Restricciones $r_{x,y} = 1$		χ^2	g.l.	Dif (χ^2)	Dif (g.l.)	Significación
Policía local	Servicios de limpieza	832,49	41	723,683	3	0,00000
Policía local	Parques y jardines	625,512	41	516,705	3	0,00000
Policía local	Gestión administrativa	1076,691	41	967,884	3	0,00000
Servicios de limpieza	Parques y jardines	370,984	41	262,177	3	0,00000
Servicios de limpieza	Gestión administrativa	646,38	41	537,573	3	0,00000
Parques y jardines	Gestión administrativa	536,187	41	427,38	3	0,00000

Fuente: Elaboración propia.

Tabla 4. Validez discriminante. Intervalos de confianza de las correlaciones estimadas

Correlaciones		Estimación de "r _i "	"Z _i "	Lim sup "Z _i "	Lim inf "Z _i "	Lim sup "r _i "	Lim inf "r _i "	Interpretación
Policía local	Servicios de limpieza	0,25	0,255	0,338	0,213	0,325	0,210	No contiene el 1
Servicios de limpieza	Parques y jardines	0,62	0,725	0,807	0,683	0,668	0,593	No contiene el 1
Parques y jardines	Gestión municipal	0,44	0,472	0,555	0,430	0,504	0,406	No contiene el 1
Policía local	Parques y jardines	0,36	0,377	0,459	0,335	0,429	0,323	No contiene el 1
Policía local	Gestión municipal	0,41	0,436	0,518	0,394	0,476	0,374	No contiene el 1
Servicios de limpieza	Gestión municipal	0,45	0,485	0,567	0,443	0,513	0,416	No contiene el 1

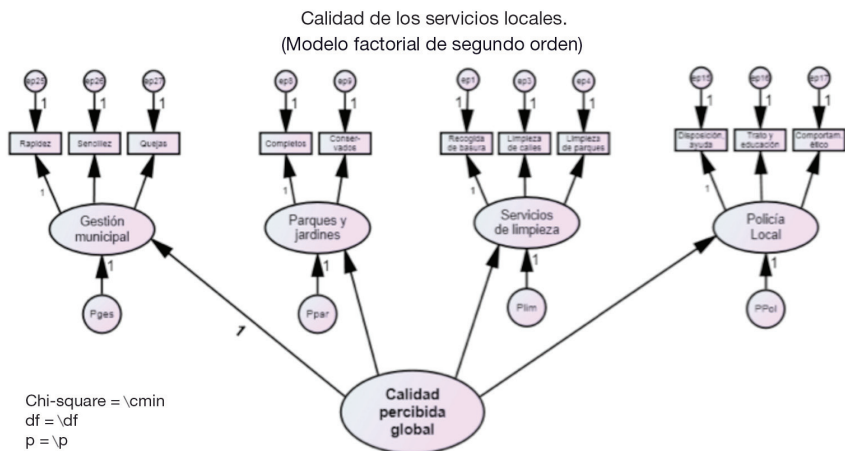
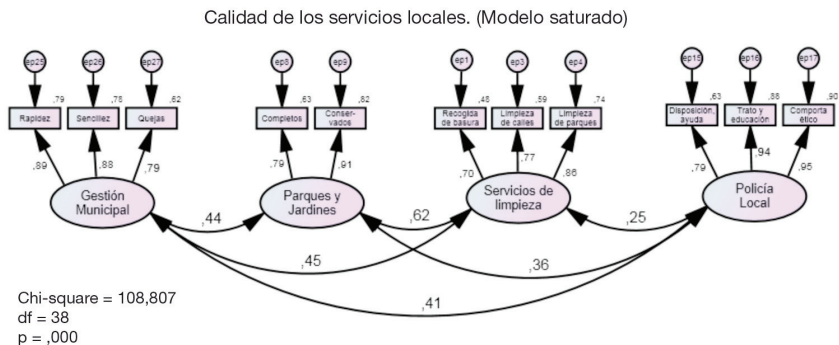
Fuente: Elaboración propia.

tres índices incrementales para determinar la pérdida de ajuste de los modelos anidados, ΔCFI , $\Delta SRMR$ y el $\Delta RMSEA$ (Chen, 2007: 501)¹⁴.

Al verificar los ajustes de ambos modelos (tabla 5) con los criterios aceptados de Hu y Bentler (1999), vemos que el modelo que contiene la calidad global percibida como factor de orden superior, que aglutina a los anteriores, cumple los criterios de ajuste de un modo excelente. De otro lado, los tres índices incrementales tampoco superan los umbrales de pérdida significativa de ajuste del modelo con calidad global respecto del saturado: $\Delta CFI = |0,007| \leq 0,01$, $\Delta SRMR = |-0,024| \leq 0,03$ y el $\Delta RMSEA = |-0,008| \leq 0,015$. Asimismo, podemos afirmar que la pérdida de ajuste detectada en el modelo de segundo orden es trivial y, por tanto, concluimos, sin demasiado riesgo, que se detecta la presencia de una idea global de calidad de los servicios locales. Así pues, la H_3 ha quedado corroborada. La calidad de servicio es un constructo de segundo orden que se origina a partir de las evaluaciones de los servicios.

14. En muestras mayores de 300 casos, no habría pérdidas significativas de ajuste cuando el índice incremental del CFI no superase en 0,01, el incremental del RMSEA en 0,015 y el incremental del SRMR en 0,03 sus valores absolutos.

Figura 3. Modelos de covarianzas saturadas y de factor de segundo orden



Fuente: Elaboración propia.

Tabla 5. Test de comparación de los modelos ajustados

	GFI	CFI	SRMR	RMSEA	Pclose _(0,06)	
Modelo covarianzas saturadas	0,967	0,982	0,033	0,057	0,626	
Modelo factor segundo orden	0,958	0,975	0,057	0,065	0,221	
Modelo de comparación	$\Delta\chi^2$	$\Delta g.l.$	p -valor	ΔCFI	$\Delta SRMR$	$\Delta RMSEA$
Modelo saturado versus modelo con factor segundo orden	28,4	2	0,0000	0,007	-0,024	-0,008

Fuente: Elaboración propia.

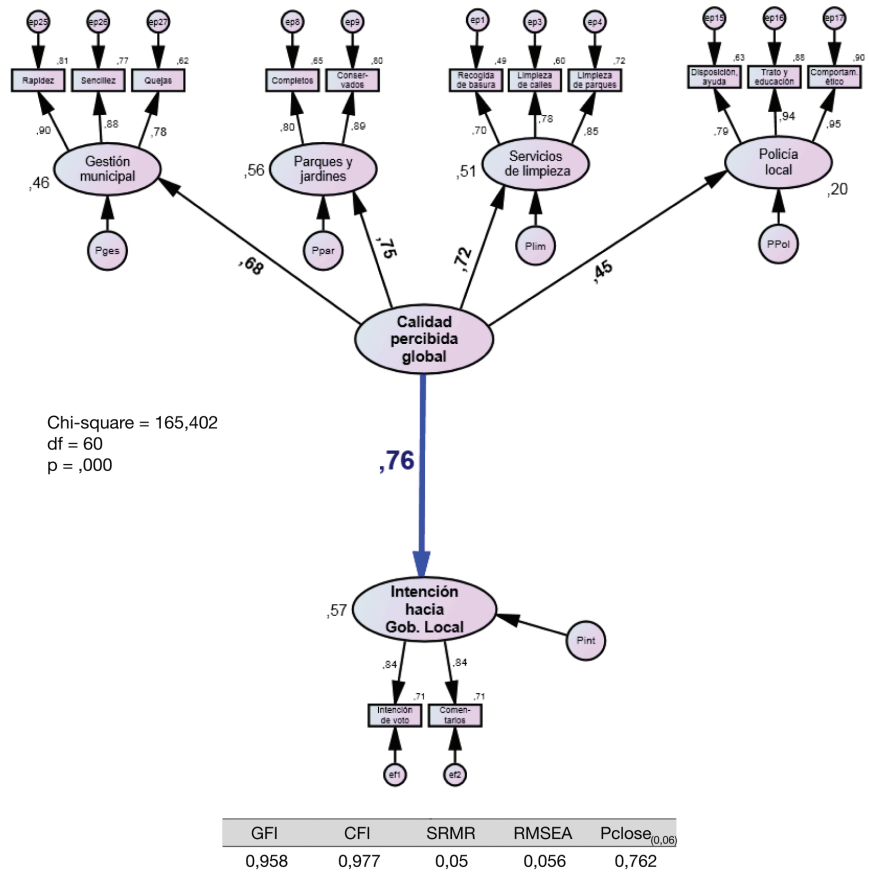
Validez nomológica. Efectos en la intención de conducta hacia el gobierno local

Para testar la vinculación teórica esperada de la calidad percibida con sus constructos, vamos a ser más exigentes que los estudios en administración que han verificado su impacto en la satisfacción del ciudadano y en su confianza con el

gobierno local (Heintzman y Marson, 2005; Van Ryzin, 2006; Roch y Poister, 2006; James, 2009).

La literatura sobre la evaluación de servicios ha demostrado que la calidad percibida está relacionada indirectamente y, por tanto, con efectos significativos, con la intención de conducta hacia el proveedor, en nuestro caso, el gobierno local (Bagozzi, 1992; Fornell et al., 1996; Oliver, 1997; Cronin et al., 2000; Harris y Goode, 2004; Johnson et al., 2001; Morgan y Hunt, 1994; Zeithalm, 2000; Johnson y Gustafsson, 2000; Keiningham, y Vavra, 2001; Gounaris, 2005; Cater y Zabkar, 2008). De esta forma, no sólo verificamos la validez nomológica del constructo, sino que también extendemos la investigación empírica de la calidad percibida en la Administración más allá

Figura 4. Impacto de la escala de calidad sobre la intención de conducta del ciudadano. Modelo estimado



Fuente: Elaboración propia.

de las fronteras de la satisfacción y de la confianza. Hemos operacionalizado la intención de conducta del ciudadano con dos indicadores: la realización de comentarios positivos o negativos sobre el gobierno local y la intención actual de voto (Zeithalm et al., 1996). El modelo siguiente (figura 4) prueba el grado en que la escala multidimensional jerárquica de la calidad percibida es capaz de relacionarse con la intención de conducta hacia el gobierno local.

El ajuste del modelo vuelve a ser bueno. El $\chi^2 = 165,402$, con 60 grados de libertad y una probabilidad de error 0,000. Los índices de ajuste seleccionados superan los estándares de buen ajuste¹⁵. La calidad percibida de los servicios está significativamente relacionada con la intención hacia el gobierno local¹⁶, con un coeficiente estándar estimado de 0,758. Los resultados no dejan lugar a dudas. La escala ha demostrado tener validez nomológica, tal y como lo espera la teoría, es decir, la calidad está relacionada con la intención de conducta hacia el proveedor. La cuarta hipótesis (H_4) ha sido corroborada por los datos. La calidad global percibida de los servicios tiene efectos lineales y positivos con la actitud intencional del ciudadano (comentarios positivos e intención de voto).

6. Discusión

La corriente de la Nueva Administración Pública apuesta por una gestión transparente, eficiente en coste y orientada al ciudadano. El problema radica en si debemos aceptar o rechazar las opiniones de la ciudadanía como medida referente de la gestión. El artículo entra en el debate y apuesta decididamente por la incorporación del juicio subjetivo del ciudadano como medida de pleno derecho en la gestión del desempeño administrativo. Se han revisado los argumentos de autores que desaconsejan una gestión gobernada por evaluaciones subjetivas, parciales en su visión, a veces resentidas por actos administrativos coercitivos o deudores del interés de clase o ideológico, y no es que muestren especial aversión por este tipo de medidas, al contrario, los mismos profesionales reconocen el valor de conceptos como *percepción ciudadana*, *satisfacción* o *confianza*, siempre que estén relegados al ámbito de estudios electorales, pero alejados del terreno de la gestión diaria (Bouckaert y Van de Walle, 2003; Roch y Poister, 2006; Howard, 2010; Micheli y Neely, 2010). Pensamos que la evaluación subjetiva tiene que incorporarse a los cuadros de mando de las administraciones públicas y que los argumentos en contra no tienen suficiente fundamento como para desbancarla. De la misma forma que los índices de satisfacción del cliente o de calidad percibida se han incorporado en las contabilidades de grandes corporaciones privadas, ese mismo movimiento debería instalarse paulatina pero inexorablemente a la contabilidad pública de la Administración. Por lo tanto, nos unimos a las voces recientes que reconocen el valor de la percepción subjetiva como referente del desempeño administrativo (Van Ryzin et al., 2008; Shingler et al., 2008; Dalehite, 2008; Morgeson III y Mithas, 2009; James, 2009; Bruning, 2010), máxime cuando en la Europa

15. Hu and Bentler (1999): GFI (> 0,95), SRMR (<0,11), CFI (> 0,95) y RMSEA (< 0,06).

16. Con un coeficiente ratio de 11,275 > 1,96.

continental, a diferencia del mundo anglosajón, hay una cierta desorientación a la hora de organizar los sistemas de medida del desempeño de la Administración local (Kuhlmann, 2010). El presente trabajo aporta al debate una herramienta y una evidencia en favor de su utilidad: la escala multidimensional jerárquica para la evaluación de la calidad percibida de los servicios de la Administración local y la capacidad ciudadana para discriminar el desempeño administrativo.

La escala multidimensional jerárquica de la calidad percibida que Dabholkar y colegas (1996) y después Brady y Cronin (2001) han aplicado con éxito a servicios del mercado libre, parece un instrumento más adecuado que la clásica escala SERVQUAL para evaluar los servicios locales. Se trata de una escala articulada en niveles. El primer nivel contiene los atributos o aspectos concretos de los servicios, el segundo delimita los servicios y el más general aglutina todos los servicios en una idea global de calidad. La forma articulada de la escala que proponemos permite construir una escala adaptada a la singularidad de cada servicio, de modo que cada servicio sólo contenga los atributos que son de aplicación a su particularidad. De esta forma, la escala aumenta considerablemente sus posibilidades de diagnóstico y, a la vez, es capaz de integrar las visiones particularizadas en una evaluación global.

Un análisis secundario realizado sobre datos de una encuesta a ciudadanos del Levante español ha validado las propiedades psicométricas de la escala, que, por otra parte, ha demostrado validez nomológica, es decir, su capacidad para enlazarse significativamente con otros conceptos que permiten realizar desarrollos teóricos nuevos en este campo. Hemos probado su relación con el concepto de la actitud intencional, lo que abre las posibilidades de ir más allá de la satisfacción o la confianza como es el paradigma intencional¹⁷.

Referencias bibliográficas

- A EVAL (2009). *Guía para la evaluación de la calidad de los servicios públicos*. Madrid: Ministerio de la Presidencia.
- ANDERSON, Rolph (1973). «Consumer Dissatisfaction: The Effect of Disconfirmed Expectancy on Perceived Product Performance». *Journal of Marketing Research*, 10 (febrero), 38-44.
- ANDERSON, James C. y GERBING, David W. (1988). «Structural Equation Modelling in practice: A Recommended Two-Step Approach». *Psychological Bulletin*, 103 (3), 411-23.
- BAGOZZI, Richard P. (1992). «The self-Regulation of Attitudes, Intentions, and behavior». *Social Psychology Quarterly*, 55, (2), 178-204.
- BAGOZZI, Richard y YI, Youjae (1988). «On the Evaluation of Structural Equation Models». *Journal of the Academy of Marketing Science*, 16 (1), 74-94.
- BITNER, Mary Jo y HUBBERT, Amy R. (1994). «Encounter Satisfaction versus Overall Satisfaction versus Quality». En: RUST, Roland T. y OLIVER, Richard L. (eds.). *Service Quality: New Directions in Theory and Practice*. Nueva York: Sage Publications, Inc.

17. En la actualidad, ya estamos trabajando en la aplicabilidad de este paradigma a la Administración local. Esperamos poderlo presentar en una próxima publicación.

- BOLLEN, Kenneth A. (1989). *Structural Equation Models with Latent Variables*. Nueva York: John Wiley and Sons. Wiley Series in Probability and Mathematical Statistics.
- BOUCKAERT, Geert y VAN DE WALLE, Steven (2003). «Comparing measures of citizen trust and user satisfaction as indicators of “good governance”: difficulties in linking trust and satisfaction indicators». *International Review of Administrative Sciences*, 69, 329-343.
- BOURGON, Jocelyne (2007). «Responsive, responsible and respected government: Towards a New Public Administration theory». *International Review of Administrative Sciences*, 73 (1), 7-26.
- BRADY, Michael K. y CRONIN, J. Joseph Jr. (2001). «Some New Thoughts on Conceptualizing Perceived Service Quality: A Hierarchical Approach». *Journal of Marketing*, 65 (julio), 34-49.
- BROWN, Karin y COULTER, Phillip B. (1983). «Subjective and Objective Measures of Police Service Delivery». *Public Administration Review*, 43 (1), 50-58.
- BROWN, Trevor (2007). «Coercion versus Choice: Citizen Evaluations of Public Service Quality across Methods of Consumption». *Public Administration Review*, 67 (3), 559-572.
- BROWN, Tom J.; CHURCHILL, Gilbert A. Jr. y PETER, J. Paul (1993). «Improving the Measurement of Service Quality». *Journal of Retailing*, 69 (1), 127-139.
- BROWNE, M.W. y CUDECK, R. (1993). «Alternative Ways of Assessing Model Fit». En: BOLLEN, K.A. y LONG, J.S. (eds.). *Testing Structural Equation Models*. Newbury Park, California: Sage.
- BRUNING, E. (2010). «A Methodological Assessment of Ten Years of Canada’s First Satisfaction Survey Research». *International Review of Administrative Sciences*, 76 (1), 85-91.
- CARMAN, James M. (1990). «Consumer Perceptions of Service Quality: An Assessment of the *Servqual* Dimensions». *Journal of Retailing*, 66 (1), 33-55.
- CARRERAS, Enrique (2000). *Revisión del modelo ECSI: Resultados España (Propuesta de un índice parsimonioso)*. Informe presentado al comité técnico del ECSI en París. Manuscrito no publicado. Ejemplar en la Asociación Española para la Calidad. Madrid, España.
- CATER, Barbara y ZABKAR, Vesna (2008). «Antecedents and consequences of commitment in marketing research services: The client’s perspective». *Industrial Marketing Management*, 30, 1-13.
- CHEN, Fang Fang (2007). «Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance». *Structural Equation Models*, 14 (3), 464-504.
- CHEN, Fang Fang; SOUSA, Karen H. y WEST, Stephen G. (2005). «Testing measurement Invariance of Second-Order Factor Models». *Structural Equation Modeling*, 12 (3), 471-492.
- CHURCHILL, Gilbert y SURPRENANT, Carol (1982). «An Investigation into Determinants on Consumer Satisfaction». *Journal of Marketing Research*, XIX (noviembre), 491-504.
- CRONBACH, Lee J. y MEEHL, Paul E. (1955). «Construct Validity in Psychological Tests». *Psychological Bulletin*, 52, 281-302.
- CRONIN, J. Joseph Jr. y TAYLOR, Steven A. (1992). «Measuring service quality: A reexamination and extension». *Journal of Marketing*, 56 (julio), 55-68.
- (1994). «SERVPERF versus SERVQUAL: Reconciling Performance Based and Perceptions-Minus-Expectations Measurement of Service Quality». *Journal of Marketing*, 58 (1), 125-131.

- CRONIN, J. Joseph Jr.; BRANDY, Michael K. y HULT, G. Tomas 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.
- DABHOLKAR, Prathiba et al. (1996). «A Measure of Service Quality for Retail Stores». *Journal of the Academy of Marketing Science*, 24 (invierno), 3-16.
- DALEHITE, Esteban G. (2008). «Determinants of Performance measurement: An Investigation into the Decision to Conduct Citizen Surveys». *Public Administration Review*, 68 (5), 891-907.
- FORNELL, Claes y LARCKER, David F. (1981). «Evaluating Structural Equation Models with Unobservable Variables and Measurement Errors». *Journal of Marketing Research*, 18 (febrero), 39-50.
- FORNELL, Claes et al. (1996). «The American Customer Satisfaction Index: Nature, Purpose, and Findings». *Journal of Marketing*, 60 (octubre), 7-18.
- GOUNARIS, Spiros P. (2005). «Trust and commitment influences on customer retention: Insights from business-to-business services». *Journal of Business Research*, 58, 126-140.
- GROVES, Robert M. et al. (2009). *Survey Methodology*. 2a ed. New Jersey: John Wiley and Sons.
- HARRIS, Lloyd C. y GOODE, Mark M.H. (2004). «The four levels of loyalty and the pivotal role of trust: a study of online service dynamics». *Journal of Retailing*, 80, 139-158.
- HEINTZMAN, Ralph y MARSON, Brian (2005). «People, Service and Trust: Is there a Public Sector service Value Chain?». *International Review of Administrative Sciences*, 71 (4), 549-575.
- HOOGLAND DEHOOG, Ruth et al. (1990). «Citizen Satisfaction with Local Governance: A Test of Individual, Jurisdictional, and City-Specific Explanations». *Journal of Politics*, 52 (3), 807-837.
- HOWARD, Cosmo (2010). «Are we being served?: A critical perspective on Canada's Citizens First satisfaction surveys». *International Review of Administrative Sciences*, 76, 65-83.
- HU, L. y BENTLER, Peter M. (1999). «Cutoff criteria for fit indexes in covariance structure analysis: Conventional Criteria versus new alternatives». *Structural Equation Modeling: A Multidisciplinary Journal*, 6, 1-55.
- HYDE, A.C. (1991). «Feedback from Customers, Clients, and Captives». *The Bureaucrat*, 20 (4), 49-53.
- JAMES, Oliver (2009). «Evaluating the Expectations Disconfirmation and Expectations Anchoring Approaches to Citizen Satisfaction With Local Public». *Journal of Public Administration Research and Theory*, 19 (1), 107-123.
- INE (Instituto Nacional de Estadística) (2008). *Cifras oficiales de población resultantes de la revisión del Padrón municipal a 1 de enero de 2008*. www.ine.es.
- JAMES, Oliver y JOHN, Peter (2007). «Public Management at the Ballot Box: Performance Information and Electoral Support for Incumbent English Local Government». *Journal of Public Administration Research and Theory*, 17, 567-580.
- JOHNSON, Michael D. y GUSTAFSSON, Anders (2000). *Improving Customer Satisfaction, Loyalty and Profit: An Integrated Measurement and Management System*. San Francisco: Jossey - Bass INC.
- JOHNSON, Michael et al. (2001). «The Evolution and Future of National Customer Satisfaction Index Models». *Journal of Economic Psychology*, 22, 217-245.
- KAMENSKY, John M. (1996). «The Role of the Reinventing Government Movement in federal Management Reform». *Public Administration Review*, 56 (3), 247-255.

- KAMPEN, Jarl K. et al. (2006). «Assessing the Relation between Satisfaction with Public Service Delivery and Trust in Government: The Impact of the Predisposition of Citizen Toward Government on Evaluations of Its Performance». *Public Performance and Management Review*, 29 (4), 387-404.
- KEININGHAM, T.L. y VAVRA, Terry (2001). *The Customer Delight Principle: Exceeding Customers' Expectations for Bottom-Line Success*. Nueva York: McGraw Hill.
- KELLY, Janet M. (2005). «The Dilemma of the Unsatisfied Customer in a Market Model of Public Administration». *Public Administration Review*, 65 (1), 76-84.
- (2007). «The Accountability Trap». *National Civic Review*, 96 (3), 46-51.
- KELLY, Janet M. y SWINDELL, David (2002). «Service Quality Variation Across Urban Space: First Steps Toward a Model of Citizen Satisfaction». *Journal of Urban Affairs*, 24 (3), 271-288.
- KETTL, Donald F. (2000). *The Global Public Management Revolution: A Report of the Transformation of the Government*. 2a ed. Washington DC: Brookings Institution.
- KLINE, Rex (1998). *Principle and Practice of Structural Equation Modeling*. Nueva York: Division of Guilford Publications.
- KUHLMANN, Sabine (2010). «Performance Measurement in European local governments: a comparative analysis of reform experiences in Great Britain, France, Sweden and Germany». *International Review of Administrative Sciences*, 76 (2), 331-345.
- LAM, Shun Yin et al. (2004). «Customer Value, Satisfaction, Loyalty, and Switching Costs: An Illustration from a Business to Business Service Context». *Journal of the Academy of Marketing Science*, 32 (3), 293-311.
- LICARI, Michael J. et al. (2005). «The Condition of Community Streets and Parks. A Comparison of Residential and Nonresidential Evaluations». *Public Administration Review*, 65 (3), 360-368.
- LYONS, William et al. (1992). *The Politics of Disatisfaction: Citizen, Services, and Urban Institutions*. Armonk, Nueva York: M. E. Sharpe.
- MACKINNON, David P. (2008). *Introduction to Statistical Mediation Analysis*. Nueva York: Lawrence Erlbaum Associates.
- MELKERS, Julia y THOMAS, John Clayton (1998). «What Do Administrators Think Citizen Think?: Administrator predictions as an Adjunct to Citizen Surveys». *Public Administration Review*, 58 (4), 327-334.
- MICHELI, Pietro y NEELY, Andy (2010) «Performance Measurement in the Public Sector in England: Searching for the Golden Thread». *Public Administration Review*, 70 (4), 591-600.
- MORGAN, Robert M. y HUNT, Shelby D. (1994). «The commitment-Trust Theory of Relationship Marketing». *Journal of Marketing*, 54 (julio), 20-38.
- MORGESON III, Forrest V. y MITHAS, Sunil (2009). «Does E-Government Measure Up to E-Business?: Comparing End User Perceptions of US Federal Government and E-Business Web Sites». *Public Administration Review*, 69 (4), 740-752.
- NGOBO, Paul Valentin (2004). «Drivers of Customers' cross-buying intentions». *European Journal of Marketing*, 38 (9/10), 1129-1157.
- NUNNALLY, Jum C. y BERSTEIN, Ira H. (1994). *Psychometric Theory*. Nueva York: McGraw Hill. Series in Psychology.
- OFIR, Chezy y SIMONON, Otamar (2007). «The Effect of Stating Expectations on Customer Satisfaction and Shopping Experience». *Journal of Marketing Research*, 64 (febrero), 164-174.
- OLIVER, Richard (1997). *Satisfaction: A Behavioral Perspective on the Consumer*. Nueva York: McGraw-Hill.

- OSHIKAWA, Sadaomi (1968). «The Theory of Cognitive Dissonance and Experimental Research». *Journal of Marketing Research*, 5 (noviembre), 429-430.
- PARASURAMAN, Anthony; ZEITHALM, Valerie A. y BERRY, Leonard L. (1988). «SERVQUAL: A Multiple-Item Scale for Measuring Consumer Perceptions of Service Quality». *Journal of Retailing*, 64 (1), 12-40.
- (1994). «Reassessment of Expectations as a Comparison Standard in Measuring Service Quality: Implications for Future Research». *Journal of Marketing*, 58 (febrero), 6-17.
- PARKS, Roger B. (1984). «Linking Objective and Subjective Measures of Performance». *Public Administration Review*, 44 (2), 118-127.
- POISTER, Theodore y STREIB, Gregory (1999). «Performance Measurement in Municipal Government: Assessing the state of the Practice». *Public Administration Review*, 59 (4), 325-335.
- ROCH, Christine H. y POISTER, Theodore H. (2006). «Citizens, Accountability, and Service Satisfaction: The Influence of Expectations». *Urban Affairs Review*, 41 (3), 292-308.
- SHINGLER, John et al. (2008). «The Importance of Subjective Data for Public Agency Performance Evaluation». *Public Administration Review*, 68 (6), 1101-1111.
- SIRDESHMUKH, Deepak et al. (2002). «Consumer Trust, Value, and Loyalty in Relational Exchanges». *Journal of Marketing*, 66 (enero), 15-37.
- SPRENG, Richard A. et al. (1996). «A Reexamination of the Determinants of Consumer Satisfaction». *Journal of Marketing*, 60 (julio), 15-32.
- STIPACK, Brian (1979). «Citizen Satisfaction with Urban Services: Potential Misuse as a Performance Indicator». *Public Administration Review*, 39 (1), 46-52.
- TEAS, Kenneth R. (1993). «Expectations, Performance Evaluation, and Consumers' Perceptions of Quality». *Journal of Marketing*, 57 (octubre), 18-34.
- (1994). «Expectations as a Comparison Standard in Measuring Service Quality: An Assessment of a Reassessment». *Journal of Marketing*, 58 (1), 132-139.
- TSE, D. y WILTON, P. (1988). «Models of Consumer Satisfaction Formation: An Extension». *Journal of Marketing Research*, XXV (mayo), 204-12.
- VAN RYZIN, Gregg G. y IMMERSWAHR, Stephen (2004). «Derived Importance Analysis of Citizen Survey Data». *Public Performance and Management Review*, 27 (4), 144-173.
- VAN RYZIN, Gregg G. (2006). «Testing The Expectancy Disconfirmation Model of Citizen Satisfaction with Local Government». *Journal of Public Administration Research and Theory*, 16, 599-611.
- (2007). «Pieces of a Puzzle: Linking Government, Citizen Satisfaction, and Trust». *Public Performance and Management Review*, 30 (4), 521-535.
- VAN RYZIN, Gregg G. et al. (2008). «Measuring Street Cleanliness: A comparison of Nueva York City's Scorecard and Results from a Citizen Survey». *Public Administration Review*, 68 (2), 295-303.
- YANG, Kaifeng y HSIEH, Jun Yi (2007). «Managerial Effectiveness of Government Performance Measurement: Testing a Middle-Range Model». *Public Administration Review*, 67 (5), 861-879.
- ZEITHALM, Valerie (2000). «Service Quality, Profitability, and the Economic Worth of Customers: What We Know and What We Need to Learn». *Journal of the Academy of Marketing Science*, 28 (1), 67-85.
- ZEITHALM, Valerie et al. (1990). *Delivering Quality Service: Balancing Customer Perceptions and Expectations*. Nueva York: The Free Press.
- (1996). «The Behavioral Consequences of Service Quality». *Journal of Marketing*, 60 (abril), 31-46.

Anexo. Cuestionario

Cuestiones relativas a la calidad percibida de los servicios

- Servicios de limpieza municipal:
 - Valoración de la limpieza en la recogida de basuras, sin dejar desperdicios por el suelo.
 - Valoración de la limpieza de las calles (sin cacas de perro ni baches).
 - Valoración de la limpieza de los parques y jardines.
- Parques y Jardines:
 - Valoración de parques completos, con zonas de paseo, deporte, niños, animales.
 - Valoración del estado de conservación de instalaciones (mobiliario, fuentes, bancos).
- Policía Local:
 - Valoración de la disposición para ayudar al ciudadano.
 - Valoración del trato y educación.
 - Valoración del comportamiento ético con el ciudadano.
- Gestión administrativa del ciudadano:
 - Valoración de la sencillez de los trámites ante el Ayuntamiento.
 - Valoración de la rapidez en la resolución.
 - Valoración de la atención a las sugerencias y quejas de la ciudadanía.

Las escalas de respuestas fueron numéricas, con nueve posiciones y anclas semánticas extremas que van desde el 1 («Muy malo») hasta el 10 («Muy bueno»).

- Intención conductual hacia el gobierno local (seleccionados de Oliver, 1997):
 - Comentarios: «Cuando Ud. habla con amigos o conocidos sobre el Ayuntamiento, ¿los comentarios suelen ser positivos o negativos?». Escala de respuesta numérica con diez posiciones, entre las cuales encontramos: 1 («Negativos»), 10 («Positivos») y 11 («No comenta»). La opción «No comenta» no se citaba para promover la respuesta.
 - Intención de voto: «Si hoy se celebraran elecciones locales, ¿con qué probabilidad votaría a favor del alcalde actual?». Escala de respuesta numérica con nueve posiciones, que van desde el 1 («Con ninguna probabilidad») hasta el 10 («Con mucha probabilidad»).
- Sexo, edad.
- Ideología declarada: «¿Diría Ud. que sus ideas suelen ser...?». Escala de respuesta semántica de siete posiciones: «Muy progresistas», «Progresistas», «Más bien progresistas», «De centro», «Más bien conservadoras», «Conservadoras» y «Muy conservadoras». Además, se abrió la posibilidad

de responder «No tiene ideología». Tanto la opción «De centro» como la de «No tiene ideología» no se citaban para motivar al entrevistado a posicionarse.

Para asegurar que el respondiente adscribe correctamente al alcalde a un partido político y después de haber expresado su intención de votar por el alcalde, se le confirma la respuesta indicando el partido, en caso contrario, se le pide que diga a qué otro partido votaría.