

Martínez, J.A. y Martínez, L. (2008). La medición de la calidad percibida en servicios deportivos; un enfoque en primera persona. Revista Internacional de Medicina y Ciencias de la Actividad Física y el Deporte vol. 8 (31) pp. 244-255
[Http://cdeporte.rediris.es/revista/revista31/artmedición96.htm](http://cdeporte.rediris.es/revista/revista31/artmedición96.htm)

LA MEDICIÓN DE LA CALIDAD PERCIBIDA EN SERVICIOS DEPORTIVOS; UN ENFOQUE EN PRIMERA PERSONA

MEASURING PERCEIVED SERVICE QUALITY IN SPORTS SERVICES; A FIRST PERSON APPROACH

Martínez, J. A.¹; Martínez, L.²;

¹ josean.martinez@upct.es, <http://www.upct.es/~beside/jose.htm>, Profesor Ayudante.

Departamento de Economía de la Empresa. Universidad Politécnica de Cartagena. España

² laura.martinez@upct.es, <http://www.upct.es/~beside/laura.htm>, Profesor Contratado Doctor.

Departamento de Economía de la Empresa. Universidad Politécnica de Cartagena. España

Recibido 1 septiembre 2008

CLASIFICACIÓN UNESCO: 5311 “Marketing”

RESUMEN

El objetivo de esta investigación es comprobar si existen diferencias entre el patrón de respuestas dado en una escala de formato libre, o enfoque en primera persona, y el realizado en las escalas comúnmente utilizadas en la medición de la calidad percibida de servicios deportivos: Likert de 1 a 5 y de 1 a 7, y diferencial semántico de -3 a +3. Este estudio ha contado con dos muestras de consumidores de centros deportivos, y combina a nivel metodológico la noción de invarianza de escala y la utilización de los modelos de ecuaciones de estructuras de covarianza. Los resultados muestran que existe distorsión estadística entre los diferentes formatos de respuesta analizados y la escala preferida por los individuos. En aras de maximizar la validez de las respuestas obtenidas, se recomienda la utilización del enfoque en primera persona.

PALABRAS CLAVES: Calidad percibida, gestión deportiva, escalas de medida, invarianza de escala.

ABSTRACT

In the framework of sports services, the aim of this research is to compare the response pattern that comes from a free-scale -or first person approach-, with the response pattern that comes from three widely used rating scales: Likert from 1 to 5 and from 1 to 7, and semantic differential from -3 to +3. This study have analysed two samples of consumers, combining the notion of scale invariance and the implementation of structural equation modelling. Our findings show that there are statistical differences between the preferred scale and the other three rating scales. With the aim to maximize validity, we recommend managers of sports services to use the first person approach.

KEYWORDS: Perceived quality, sport Management, rating scales, scale invariance.

1. INTRODUCCIÓN

La *gestión de la calidad* es ya un concepto universalmente conocido y aplicado en prácticamente todos los ámbitos de la dirección de organizaciones y empresas de servicios. En los últimos años, la gestión de servicios deportivos, aunque con un cierto retraso, no ha sido ajena a esta tendencia, y ha ido incorporando también esta filosofía a los procesos de dirección.

Un aspecto fundamental en la implementación de cualquier tipo de sistema que pretenda mejorar la eficiencia y eficacia de los procesos y conseguir la satisfacción y retención del cliente, es el de establecer un sistema de medición que permita escuchar la voz del consumidor, siendo éste, en última instancia, el que otorga una valoración subjetiva sobre la calidad del servicio recibido. De este modo, las organizaciones pueden utilizar esa medición para evaluar si el esfuerzo realizado a nivel de gestión, se corresponde con las evaluaciones de sus clientes. Éste es un principio básico de sistemas de gestión de la calidad como, por ejemplo, la Norma ISO 9000, o los modelos de autoevaluación EFQM o CAF.

Desde los años 80, se han realizado numerosas aportaciones en la literatura especializada sobre la utilización de instrumentos de medida para evaluar la calidad percibida, proponiendo modelos o escalas que han sido debatidos y cuestionados a nivel teórico, metodológico y empírico. Ejemplos de estas aportaciones son los trabajos de Grönroos (1984), Cronin y Taylor (1992), Parasuraman, Zeithalm y Berry (1985; 1988), Teas (1993) o Brady y Cronin (2001). Por su parte, en el ámbito de la gestión deportiva, destacan las investigaciones de Kim y Kim (1995), Ko y Pastore (2004, 2005) o Morales, Hernández-Mendo y Blanco (2005).

Sin embargo, ninguna de estas aportaciones cuestiona un aspecto primordial en la medición de la calidad, como es la utilización de un formato de respuesta óptimo que maximice la validez de las respuestas obtenidas. El motivo fundamental de no profundizar en este tema reside en que esta cuestión

forma parte de otro debate, mucho más profuso si cabe, que aproximadamente desde los años 30 del siglo pasado se protagoniza en las ciencias sociales en general, y particularmente en la psicología y sociología.

En este escenario de discusiones científicas, la mayoría de los autores que investigan sobre calidad percibida, ya sea por prudencia algunos y por rutina otros, utilizan generalizadamente las escalas Likert de 1 a 5 o de 1 a 7, apoyándose sobre todo en trabajos como los de Lissitz y Green (1975) o Cox (1980). Del mismo modo, las escalas diferencial semántico también son muy populares, gracias principalmente a la distinción entre emociones positivas y negativas de Kahneman y Tversky (1979) en su teoría de las perspectivas. No obstante, trabajos como los de Hofmans, Theuns y Mairesse (2007) o Weng (2004), muestran las numerosas contradicciones que se han derivado de los estudios empíricos acerca del número adecuado de alternativas de respuesta, la elección de etiquetas verbales para esas alternativas, la fiabilidad y validez de escalas de diferente rango, el efecto de presentar una respuesta neutra, o la preferencia de los consumidores entre diferentes formatos de respuesta.

En aras de mejorar la validez de las mediciones, teóricamente sería más adecuado que los consumidores respondieran de la forma que les resultase preferida, es decir, no restringiendo las respuestas a una escala dada por el investigador, sino siendo el propio individuo quien respondiera en la “escala” que más se ajustase a sus preferencias, al contexto determinado de la pregunta y respuesta, y que minimizara el esfuerzo o coste psicológico (Ferrando, 2003; Weng y Cheng, 2000). Por tanto, el hecho de que el individuo definiera su respuesta sin restricciones de categorización sería, en terminología relativista, una aproximación en “primera persona” a la medida de variables psicológicas (Kilpatrick y Cantril, 1960), en contraposición con la llamada perspectiva en “tercera persona”, asociada al positivismo, que es la forma habitual de proceder en la tradicionalmente etiquetada como investigación cuantitativa. Sin entrar, no obstante, en disquisiciones filosóficas, parece evidente que esa aproximación en primera persona incrementa la validez de las respuestas, por lo que es una opción muy deseable para los investigadores, y más teniendo en cuenta que la evidencia empírica ha demostrado que los individuos utilizan un rango determinado de respuestas (Ferrando 2003), o que directamente prefieren dar sus respuestas en escalas de 1 a 10 (Preston y Colman, 2000). Esto indica que puede existir homogeneidad por parte de los individuos en la forma de contestar. No obstante, y tal y como recomiendan Ferrando (2003) o Hofmans, Theuns y Van Acker (en prensa), es conveniente analizar cada contexto de investigación por separado, abogando por la especificidad de los estudios.

Además del incremento de la validez por ausencia de categorización, un formato libre de respuesta permite que las etiquetas verbales comúnmente utilizadas en las escalas ordinales no produzcan interacción con la respuesta, es decir, no sean fuente de sesgo. Y es que, por ejemplo, las diferencias culturales pueden producir divergencias en los patrones de respuesta (Hofmans et al., en prensa). Incluso algunos individuos pueden considerar, por

ejemplo, “muy bueno” como el punto final de una escala, mientras que otros no (Sarís y Gallhofer, 2007), por lo que el estilo de respuesta varía incluso en función de la personalidad del encuestado (Javaras y Ripley, 2007). Dadas estas circunstancias, por tanto, se podría obtener una medida numérica de la percepción de la calidad del consumidor de servicios deportivos optimizando la validez de la respuesta al utilizar una escala de formato libre.

El objetivo de esta investigación es comprobar si estadísticamente existen diferencias entre el patrón de respuestas dado en una escala de formato libre, y el realizado en las escalas comúnmente utilizadas: Likert de 1 a 5 y de 1 a 7 y diferencial semántico de -3 a +3. De este modo, si esa diferencia es relevante, los gestores deportivos deberían replantearse la forma en la que evalúan la calidad del servicio. Este estudio ha contado con dos muestras de consumidores de servicios deportivos (la segunda de ellas de replicación), y combina a nivel metodológico la noción de invarianza de escala y la utilización de los modelos de ecuaciones de estructuras de covarianza.

2. METODOLOGÍA

2.1. INVARIANZA DE ESCALA

En psicofísica, la invarianza de escala es conocida como una característica de los objetos que no cambia si la longitud de la escala es multiplicada por un factor constante. Por ejemplo, dada la función polinómica $f(x) = ax^k$, donde a y k son constantes, entonces $f(cx) = c^k ax^k = c^k f(x)$, donde c es una constante. Es decir, escalando el argumento de la función por un factor constante c , se produce un re-escalamiento de la función por un factor constante c^k .

Los trabajos de Stevens (1951) mostraron que la respuesta a un estímulo $f(x)$ es una función de potencia de la intensidad percibida x , por lo que se cumple que $f(x) = ax^k$. Si suponemos $k=1$ (Hofmans et al., 2007), entonces tomando logaritmos: $\ln(f(x)) = \ln(a) + \ln(x)$. Si la respuesta al estímulo se da en la escala que el individuo mejor relaciona con la intensidad que percibe, es lógico suponer que $a=1$, por lo que la relación es completamente idéntica entre la respuesta y la intensidad percibida. Si existe un cambio de escala, entonces: $\ln(f(cx)) = \ln(c) + \ln(a) + \ln(x)$. Despejando c , se obtiene el factor de re-escalamiento. Por ejemplo, para una escala preferida de 0 a 10, si se mide el estímulo en una escala de 1 a 5, el valor de $c=0.4$. Es decir, si el estímulo original tiene una intensidad de 5, le correspondería un valor re-escalado de 2. Hay que resaltar que como los orígenes de las escalas son distintos, el valor re-escalado pertenece a una escala de 5 categorías pero con origen en cero (0,1,2,3,4), que fácilmente se corresponde con el valor 3 en la escala de 1 a 5.

Como puede intuirse, el simple hecho de transformar los valores de 0 a 10 en una escala de 1 a 5 implica un sesgo por categorización, ya que a ciertos valores de la primera escala le corresponderían valores decimales de la segunda, que deben ajustarse a los valores enteros disponibles (ver ilustración en Cox (1980)). La cuestión es si esas distorsiones que se producen en los valores de respuesta son lo suficientemente relevantes para significar un error a tener en cuenta.

Igualmente, dado que la varianza S^2 de una distribución muestral de n datos es una función cuadrática, es sencillo comprobar como $g\left(\sum_{i=1}^n f(x)\right) = S^2$, donde $f(x) = ax^k$, $x = (x_i - \bar{x})$, $k=2$ y $a=1/n$. Por tanto, si existe invarianza de escala, la varianza re-escalada sería $f(cx) = c^k ax^k$, es decir, $g\left(\sum_{i=1}^n f(cx)\right) = c^2 S^2$, siendo S^2 la varianza de la escala original.

Una vez definida la invarianza de escala, a nivel operativo es más accesible trabajar transformando todas las respuestas de escalas distintas en una única escala universal en el intervalo $[0,1]$ (Cohen, Cohen, Aiken y West, 1999). De este modo, se puede establecer una comparación directa entre todas las respuestas.

2.2. MODELOS DE ECUACIONES ESTRUCTURALES

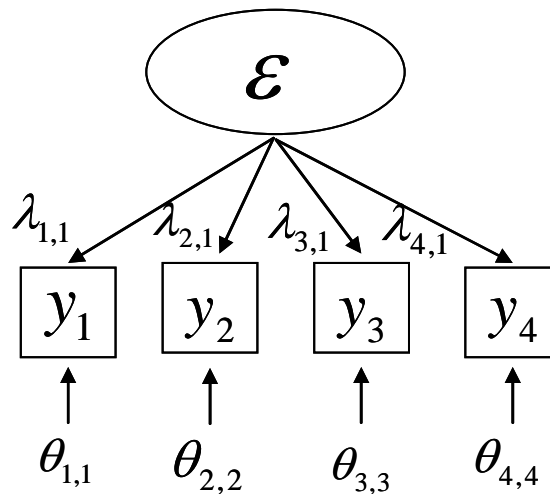
Los modelos lineales de ecuaciones de estructuras de covarianza son ampliamente utilizados en ciencias sociales, permitiendo el estudio de relaciones entre variables latentes (no observables), y entre éstas e indicadores (respuestas observables). El modelo reflectivo más sencillo que puede plantearse es el que relaciona una variable latente con un indicador, cuya ecuación es: $y = \lambda\varepsilon + \theta$, siendo y el valor del indicador observable, ε el valor de la variable latente, λ el coeficiente que relaciona ambos valores, y θ un término de error. Rápidamente se ve la analogía de esta ecuación con las definidas en la sección anterior. En este caso, el término de error se asume que sigue una distribución normal con media cero, por lo que no influye en el valor medio de los valores de respuesta, aunque sí incrementa la varianza observable de esas respuestas.

La ecuación anterior puede expresarse en términos de covarianza de la siguiente manera: $Var(y) = \lambda^2 Var(\varepsilon) + Var(\theta)$, asumiendo independencia en la parte derecha de la ecuación.

El modelo que proponemos en esta investigación es una extensión de esas ecuaciones anteriores. En este caso, la variable latente hace referencia a

la percepción del estímulo, es decir, la calidad percibida, que se manifiesta a través de cuatro indicadores observables. El primero de ellos, y_1 , representa la escala preferida por el individuo. Los tres restantes, y_2 , y_3 e y_4 , representan las respuestas a ese estímulo en escala diferencial semántico de -3 a +3, y en escalas Likert de 1 a 5 y a 1 a 7 respectivamente. De esta forma, el modelo quedaría ilustrado según la Figura 1.

Figura 1
Modelo de medida de la calidad percibida con las cuatro escalas consideradas



Dado que se cumplen las condiciones de identificación, se pueden testar diferentes modelos atendiendo a ciertas restricciones en los parámetros a estimar. Seguiremos la filosofía de ajuste exacto, defendida por Hayduk (1996) y Hayduk y Glaser (2000), por lo que los modelos se considerarán ajustados en base únicamente al test de la chi-cuadrado.

Partimos de la proposición de que la relación entre la respuesta y la intensidad percibida es completamente idéntica ($\alpha=1$, o en el caso de la notación de ecuaciones estructurales $\lambda_{i,1}=1$). Como todos los datos están transformados a una escala universal [0,1], si existe invarianza de escala, el resto de parámetros lambda también deberían ser 1 (modelo de indicadores tau-equivalente). Asimismo, asumimos que puede existir error de medida en la respuesta de los consumidores, el cual consideramos que tiene una varianza del orden del 15% de la varianza observada (Martínez y Martínez, 2008). En este caso una condición aún más restrictiva para que se cumpla la invarianza de escala es que las varianzas de error sean iguales para los cuatro indicadores (modelo de indicadores paralelos).

2.3. RECOGIDA DE DATOS

Se recogió una muestra aleatoria de 116 consumidores de uno de los centros deportivos municipales (el más importante) de la ciudad de Cartagena. Esa muestra constituye alrededor del 6% de usuarios de esos servicios. El factor de imprecisión sobre la escala de medida- *FIEM*- (Martínez y Martínez, 2008) fue del 4,5%, factor que se tomó como criterio para la determinación de un tamaño de muestra mínimo (<5%). La muestra estaba compuesta por un 69% de hombres y un 31% de mujeres, siendo la media de experiencia en el servicio de 31 meses. La recogida de datos se realizó durante la primavera de 2008.

A los encuestados se les pidió que expresaran su percepción de calidad del servicio deportivo, a través de un término lingüístico y a través de un valor numérico. Cuando el encuestado expresaba ese valor numérico se le pedía que indicara la escala en la que ese valor cobraba sentido. Tras dos preguntas que no versaban sobre calidad percibida, se les volvía a preguntar sobre su percepción de calidad del servicio utilizando las escalas categóricas descritas anteriormente. Entonces el encuestador afirmaba: *“Usted a dicho que su percepción de calidad era XXX (y utilizaba el término lingüístico que había dicho el encuestado), ¿cómo lo representaría en las siguientes escalas?”*. De este modo, separando las preguntas sobre calidad en dos bloques y con la comentada intervención del encuestador, se perseguía minimizar la posibilidad de que existiera dependencia entre las respuestas numéricas, lo que convertiría el modelo en no testable, al no cumplir las condiciones de identificación.

Finalmente, un 52% de los encuestados escogieron como escala preferida para dar su respuesta sobre la percepción de calidad una escala de 1 a 10, mientras que un 48% lo hizo en una escala de 0 a 10.

3. RESULTADOS

Para analizar la invarianza de escala utilizamos el programa LISREL 8.80 (Jöreskog y Sörbom, 2006), y el método de estimación de máxima verosimilitud robusto, ya que los datos se alejaban moderadamente de la normalidad multivariante. Consideramos todas las variables de entrada como continuas. Hemos seguido un procedimiento de test basado en la comparación de diferentes modelos anidados, comparando el ajuste de esos modelos a través de la diferencias de la chi-cuadrado escalada de Satorra-Bentler (Satorra y Bentler, 1999).

La secuencia de modelos anidados se describe en la Tabla 1. Como puede contemplarse, el modelo base y el de ítems tau-equivalente se ajustan, pero no el modelo de indicadores paralelos (M_3). Sin embargo, M_2 tiene un ajuste comparativamente peor que M_1 , ya que $\Delta SB\chi^2(gI)$ es significativa ($p < 0.05$). No obstante, M_2 es un modelo estadísticamente aceptable, aunque en la frontera de la significación.

Tabla 1
Modelos estadísticos

Modelo	Restricciones	$SB\chi^2 (gl)$	p valor	$\Delta SB\chi^2 (gl)$	p valor
M_1	Modelo base	4.34 (3)	0.23		
M_2	Modelo de items tau-equivalentes	12.55 (6)	0.05	$M_2 - M_1 = 21.94 (3)$	0.00
M_3	Modelo de items paralelos	19.75 (9)	0.02		
Replicación					
Mr_1	Modelo base	19.41 (3)	0.00		
Mr_2	Modelo de items tau-equivalentes	36.17(6)	0.00		
Mr_3	Modelo de items paralelos	53.97 (9)	0.00		

El estudio fue replicado en otra muestra aleatoria de otro centro deportivo municipal (el segundo más importante), con el fin de contrastar la consistencia de los resultados en un contexto diferente, es decir, ante lo que Hitchcock (2002) llama otras situaciones test. Asumimos que puede existir heterogeneidad a priori, y por tanto, diferenciación en el test, porque este centro es de gestión compartida con una empresa privada. Por tanto, partimos de la hipótesis de que la calidad percibida puede ser diferente en este segundo centro. Dada esta situación, creemos conveniente estudiar los resultados del primer estudio se confirman con esta segunda muestra.

De nuevo se estableció como criterio para obtener el tamaño de muestra el obtener un *FIEM* menor del 5%. En este caso se obtuvieron 98 cuestionarios válidos; un 60% de los encuestados eran hombres y un 40% mujeres, siendo la media de experiencia en el servicio de 40 meses.

En esta segunda muestra, un 78% prefirió la escala de 1 a 10 frente al 22% que prefirió la escala de 0 a 10.

Se siguió un procedimiento análogo al del estudio 1 en cuanto a la secuencia de modelos estadísticos (Tabla 1). En este caso, ni siquiera el modelo base Mr_1 puede considerarse correctamente especificado, por lo que tampoco tienen sentido las comparaciones relativas entre modelos (Yuan y Bentler, 2004).

De este modo, la replicación ofrece resultados diferentes a los del primer estudio. En esta segunda muestra no se ajusta el modelo base, lo que indica que los ítems no son condicionalmente independientes, por lo que no conforman una estructura básica de medición de la calidad percibida.

4. CONCLUSIONES

De esta investigación pueden desprenderse una serie de conclusiones de relevancia para la medición de la calidad de servicios deportivos.

En primer lugar, los consumidores de servicios deportivos prefieren expresar su percepción de la calidad del servicio en escalas de 1 a 10 o de 0 a 10, esta última con la posibilidad de dar valores decimales. La escala de 1 a 10 se muestra como claramente preferida en el segundo estudio, coincidiendo con los resultados de Preston y Colman (2000). Esto indica que, aunque las actitudes tengan teóricamente un carácter latente y continuo, existe una gran unanimidad en la expresión o manifestación de las mismas en este tipo de escalas, lo que muestra que existe un marco de referencia común sobre el cual los consumidores hacen sus evaluaciones. Posteriores estudios deberán poner atención sobre si realmente no existe divergencia entre ambos formatos de respuesta. Y es que la distinción entre 1 y 0 en el límite inferior podría ser debida a la inexistencia de percepciones de calidad extremadamente negativas. De hecho, sólo un individuo respondió con el peor valor posible (y lo hizo utilizando el 0).

En segundo lugar, la evidencia empírica muestra que efectivamente existe distorsión estadística entre los diferentes formatos de respuesta analizados y la escala preferida por los individuos. Aunque los resultados del primer estudio indican que puede existir invarianza de escala en su versión menos restrictiva (modelo de ítems paralelos, que está justo en el límite de la significación estadística: 0.05), la replicación da unos resultados mucho más claros, donde incluso no existe ajuste del modelo base. Este último factor contradice el principio de independencia condicional, el cual debe ser necesario en su cumplimiento en modelos de medida como el que se han planteado.

Dados estos resultados, consideramos más prudente recomendar la utilización de escalas libres, es decir, de medir la calidad percibida en “primera persona”, sin obligar al encuestado a contestar en una escala categórica constreñida en 5 o 7 categorías de respuesta, ya sea Likert o diferencial semántico, y que distorsiona su respuesta de forma estadísticamente significativa. En aras de maximizar la validez de las respuestas obtenidas, creemos que ésta es una recomendación a tener muy en cuenta.

Finalmente, la principal limitación de este estudio reside en la posibilidad de que existan modelos competitivos de dependencia entre indicadores. Aunque hemos tratado de minimizar ese posible riesgo en el diseño del cuestionario, admitimos que es plausible que se pueda producir esa dependencia. El hecho de que en la segunda muestra no se ajuste el modelo base, podría ser indicativo de ese hecho. Para poder testar modelos de dependencia sería necesario incrementar el número de variables observables utilizadas, con el fin de poder identificar el modelo. Si esa dependencia existe, los resultados de nuestro estudio quedarían sesgados. No obstante, insistimos en que la forma bien explicitada de realizar la encuesta a cada individuo, minimiza ostensiblemente ese riesgo.

REFERENCIAS

- 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.
- Cohen, P., Cohen, J., Aiken, L., y West, S. (1999). The problem of units and the circumstance for POMP. *Multivariate Behavioral Research*, 34 (3), 315-346.
- Cox, E. P. (1980). The optimal number of response alternatives for a scale: a review. *Journal of Marketing Research*, 17, 407-422
- Cronin, J.J., y Taylor, S. (1992). Measuring service quality: a reexamination and extension. *Journal of Marketing*, 56, 55–68.
- Ferrando, P.J. (2003). A Kernel density analysis of continuous typical-response scales. *Educational and Psychological Measurement*, 63, 809-824
- Grönroos, C. (1984). A service quality model and its marketing implications. *European Journal of Marketing*, 18 (4), 36–44.
- 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 analysis, and other serious fun. *Structural Equation Modeling*, 7, 1–35.
- Hitchcock, C. (2002). *Probabilistic causation*. Stanford Encyclopedia of Philosophy.
- Hofmans J., Theuns P., Baekelandt S., Mairesse O., Schillewaert N., y Cools W. (2007). Bias and changes in perceived intensity of verbal qualifiers effected by scale orientation. *Survey Research Methods*, 1 (2), 97 -108.
- Hofmans, J., Theuns, P. y Van Acker, F. (En prensa). Combining quality and quantity. A psychometric evaluation of the self-anchoring scale. *Quality & Quantity*.
- Hofmans, J., Theuns, P., y Mairesse, O. (2007). On the impact of the number of response categories on linearity and sensitivity of 'Self Anchoring Scales'. A Functional Measurement approach. *Methodology*, 3, (4), 160-169.
- Javaras, K. N. y Ripley, B. D. (2007). An "unfolding" latent variable model for likert attitude data: drawing inferences adjusted for response style, *Journal of the American Statistical Association*, 102 (478), 454-463
- Jöreskog, K. y Sörbom, D. (2006). *LISREL 8.80*. Scientific Software International, Inc

- Kahneman, D., y Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica*, 47 (2), 263- 291.
- Kilpatrick, F. P., y Cantril, H. (1960). Self-anchoring scaling: A measure of individuals' unique reality worlds. *Journal of Individual Psychology*, 16, 158-173.
- Kim, D., y Kim, S.Y. (1995). QUESC: an instrument for assessing service quality in sports and leisure centres. *Managing Leisure* 1, 77–89.
- Ko, Y. J., y Pastore, D. L. (2004). Current issues and conceptualizations of service in the recreational sport industry. *Sport Marketing Quarterly*, 13 (3), 159–167.
- Ko, Y.J., y Pastore, D. L. (2005). A hierarchical model of service quality for the recreational sport industry. *Sport Marketing Quarterly*, 14 (2), 84–97.
- Lissitz, R. W. y Green, S. B. (1975). Effect of the number of scale points on reliability: A Monte Carlo approach. *Journal of Applied Psychology*, 60, 10-13
- Martínez, J. A. y Martínez, L. (2008). Determinación de la máxima varianza para el cálculo del Factor de Imprecisión sobre la Escala de Medida, y extensión a diferentes tipos de muestreo. *Psicothema*, 20 (2), 305-310.
- Martínez, L. y Martínez, J. A. (2008). Developing a multidimensional and hierarchical service quality model for the travel agency industry, *Tourism Management*, 29 (4), 706-720
- Morales, V., Hernández-Mendo, A., y Blanco, A. (2005). Evaluación de la calidad en los programas de actividad física. *Psicothema*, 17 (2), 311-317
- 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., (1985). A conceptual model of service quality and its implications for future research. *Journal of Marketing*, 49, 35–48.
- Preston, C. C., y Colman, A. M. (2000). Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, 104, 1-15
- Saris, W. E. y Gallhofer, I. (2007). Estimation of the effects of measurement characteristics on the quality of survey questions. *Survey Research Methods*, 1 (1), 29-43
- Satorra, A. y Bentler, P. M. (1999). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, 66,. 507-514.
- Stevens, S. S. (1951). Mathematics, measurement and psychophysics. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley.
- Teas, R. (1993). Expectations, performance evaluation, and consumer's perceptions of quality. *Journal of Marketing* 57, 18–34.
- Weng, L. (2004). Impact of the number of response categories and anchor labels on coefficient alpha and test-retest reliability. *Educational and Psychological Measurement*, 64, 956-972.
- Weng, L.-J., y Cheng, C.-P. (2000). Effects of response order on Likert-type scales. *Educational and Psychological Measurement*, 60 (6), 908–924

Yuan, K. H., y Bentler, P. M. (2004). On chi-square difference and z-tests in mean and covariance structure analysis when the base model is misspecified. *Educational and Psychological Measurement*, 64, 737–757.

[Rev.int.med.cienc.act.fis.deporte](#) - vol. 8 - número 31 - septiembre 2008 - ISSN: 1577-0354