REVISTA INTERNACIONAL DE CIENCIAS DEL DEPORTE
International Journal of Sport Science

doi:10.5232/ricyde2009.01502

International Journal of Sport Science
VOLUMEN V - AÑO V
Páginas:17-35 ISSN:1885-3137
Nº 15 - Abril - 2009

Rev. int. cienc. deporte

Estudio de la invarianza de escala mediante el método de cálculo integral en la medición de la calidad percibida de los servicios deportivos.

Analysing scale invariance through integral calculus when measuring perceived quality in sports services.

José Antonio Martínez García

Facultad de Ciencias de la Empresa
Universidad Politécnica de Cartagena, Murcia (España)

Resumen

Esta investigación presenta un nuevo método para el estudio de la invarianza de escala que complementa otros métodos existentes, lo que contribuye a realizar un análisis ecléctico y multifocal de un problema importante en la investigación de marketing, y en particular en la investigación de servicios deportivos. Este método está basado en la utilización del cálculo integral y tiene una sencilla interpretación geométrica. Se describen y comparan varios procedimientos para testar la invarianza de escala, y se realiza un re-análisis de la investigación de Martínez y Martínez (2008b) sobre la percepción de calidad del consumidor de servicios deportivos. Los resultados muestran cómo existen diferencias sobre las conclusiones originales de estos autores. De este modo, las escalas de siete opciones de respuesta sí son invariantes, mientras que la de cinco opciones no lo son. Finalmente, se discuten las bondades y las limitaciones del método integral, abogando por la triangulación estadística para dar robustez a los resultados empíricos.

Palabras clave: invarianza de escala, calidad percibida, servicios deportivos, cálculo integral.

Abstract

This research introduces a new method to analyse scale invariance, which overcomes some shortcomings of other procedures. Under an eclectic perspective, this method must help to provide insights in the marketing research discipline, and specifically in the sports service management. The method is grounded on the use of definite integrals to compute the area between two functions. In addition, several procedures for testing scale invariance are depicted and compared. An empirical application is achieved by re-analysing the study of Martínez & Martínez (2008b) on perceived quality in sports services. Results shows that misleading conclusions were derived from the original study of those authors. Finally, advantages and shortcomings of the new method are discussed.

Key words: scale invariance, perceived quality, sports services, definite integrals.

Correspondencia/correspondence: José Antonio Martínez García
Departamento de Economía de la Empresa. Universidad Politécnica de Cartagena.
Facultad de Ciencias de la Empresa. Paseo Alfonso XIII, 50. 30203. Cartagena. España.
E-mail: josean.martinez@upct.es

Recibido el 10 de octubre 2008; Aceptado el 12 de diciembre de 2008

Introducción

La preocupación por la conceptualización y medición de la calidad percibida en servicios deportivos ha alcanzado ya un nivel similar a la de otros sectores y ámbitos económicos. Prueba de ello son las continuas aportaciones que en los últimos años se han realizado en las revistas especializadas, como por ejemplo la de Tsitskari, Tsiotras y Tsiotras (2006), donde se revisan y debaten algunas de las contribuciones más relevantes en esta disciplina acerca de la propuesta de modelos de medición en gestión deportiva.

Recientemente, Martínez y Martínez (2008b) contribuyen novedosamente al estudio de este tema, discutiendo sobre la necesidad de estudiar cuál es el formato de respuesta óptimo que maximice la validez o utilidad de las contestaciones obtenidas. Así, abogan por la utilización de un formato libre de respuesta o de escalas ordinales que sean invariantes. Para ello, proponen estudiar la invarianza de escala a través de modelos de ecuaciones estructurales. En su aplicación empírica, encuentran que los consumidores prefieren responder en un formato de respuesta de 1 a 10 o de 0 a 10, y que las escalas Likert de 1 a 5 y de 1 a 7, y la diferencial semántico de -3 a +3, que son las que testan en su estudio, no son invariantes, por lo que no deben ser utilizadas en la investigación sobre calidad percibida en servicios deportivos, ya que ello sesgaría los resultados. Si las escalas de medida no son invariantes, por tanto, se produce una distorsión de los resultados estadísticos, ya que el instrumento de medida (en este caso la escala en cuestión) interaccionaría con la medición, lo que haría que ésta dependiese de aquélla. Es decir, sobre un mismo fenómeno a estudiar, instrumentos de medida distintos producirían distintos resultados, en el caso de que esos instrumentos no fueran invariantes.

Sin embargo, y dado que los modelos de ecuaciones estructurales tienen ciertas limitaciones en su aplicación, Martínez y Ruiz (2008) desarrollan un test no paramétrico basado en entropía y dinámica simbólica para analizar la invarianza de escala (*D*-test), que es bastante más flexible y menos exigente en cuanto a la demanda de asunciones sobre los datos. No obstante, el *D*-test también tiene algunos inconvenientes, como la necesidad de un tamaño de muestra mínimo y la no representación de intervalos de confianza.

En esta investigación se presenta una nueva, heurística y simple forma de estudiar la invarianza de escala basada en la diferencia de áreas entre dos funciones, y que cubre alguna de las limitaciones de otros métodos. En consonancia con la perspectiva de Tukey (1977) sobre la necesidad de establecer una multiplicidad de enfoques sobre el análisis de datos, con el fin de que esa pluralidad metodológica enriquezca las conclusiones derivadas, se estima conveniente realizar un re-análisis del estudio empírico de Martínez y Martínez (2008b).

Para ello, este trabajo resume los procedimientos más relevantes para estudiar la invarianza de escala, desarrollando además esta nueva contribución. Tras indicar las ventajas e inconvenientes de cada enfoque, se vuelven a analizar los datos del estudio empírico de Martínez y Martínez (2008b) utilizando cada uno de esos métodos. Por tanto, la aportación de esta investigación a la gestión deportiva es doble: (1) proponer un nuevo método para el estudio de la invarianza de escala; (2) analizar si existe invarianza de escala para medir la calidad percibida aplicando un enfoque multimétodo.

La invarianza de escala

Martínez y Martínez (2008b) definen la invarianza de escala 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 .

Igualmente, dado que la varianza S^2 de una distribución muestral de *n* datos es una función cuadrática, es sencillo comprobar cómo $g(\sum_{i=1}^{n} f(x)) = S^2$, donde $f(x) = ax^k$, $x = (x_i - \overline{x})$, 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, $g(\sum_{i=1}^{n} f(cx)) = c^2 S^2$, 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.

Formas de analizar la invarianza de escala

<u>Comparación directa</u>

La opción más simple es realizar una comparación directa entre las medias y las varianzas de las dos escalas (la escala libre (A) y otra escala (B) normalizadas al intervalo [0,1]). Si ambos estadísticos no difieren, entonces favorecería la hipótesis de que existe invarianza de escala.

La limitación de esta perspectiva es que las medias y las varianzas podrían ser estadísticamente iguales, pero el patrón de respuestas de cada individuo podría no ser invariante.

En el caso de la media, se podría asumir que si ambas medias son iguales, el valor esperado de todas las desviaciones individuales de la invarianza de escala es cero. Consecuentemente, se podrían considerar esas desviaciones como un error de categorización con media cero. La varianza de ese error distorsionaría la varianza de la escala B, por lo que la hipótesis de igualdad de varianzas se rechazaría con mayor asiduidad. Sin embargo, asumir que este error de categorización es una variable aleatoria con media cero puede no ser una asunción demasiado realista, porque implicaría que los individuos serían siempre capaces de mapear mentalmente cada valor de la escala A con la categoría adecuada de la escala B (el valor íntegro más cercano). Además, esta asunción requeriría que la distribución de respuestas de A fuera uniforme, o que fuera perfectamente simétrica para poder así balancear los errores de categorización positivos y negativos. Este último hecho es ciertamente difícil de asumir en la práctica.

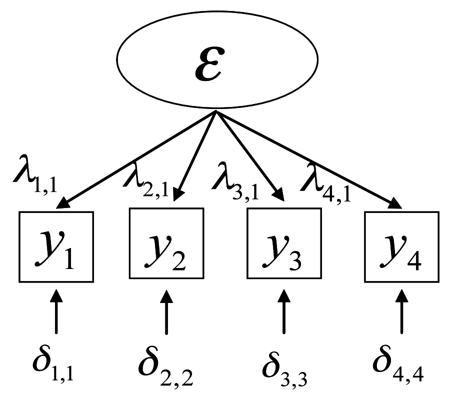
Por tanto, usando este método es difícil discernir si la hipótesis de invarianza de escala se cumple simplemente por azar.

Modelos de ecuaciones estructurales

Los modelos lineales de ecuaciones de estructura 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). Como indican Martínez y Martínez (2008b), el modelo reflectivo más sencillo que puede plantearse es el que relaciona una variable latente con un indicador, cuya ecuación es: $y = \lambda\varepsilon + \delta$, siendo y el valor del indicador observable, ε el valor de la variable latente, λ el coeficiente que relaciona ambos valores, y δ 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. De este modo, una de las ventajas de utilizar los modelos de ecuaciones estructurales es el poder considerar el error de medición de las variables latentes.

La ecuación anterior puede expresarse en términos de covarianza de la siguiente manera: $Var(y) = \lambda^2 Var(\varepsilon) + Var(\delta)$, asumiendo independencia en la parte derecha de la ecuación. La invarianza de escala puede ser estudiada utilizando cuatro escalas (la escala libre: y_1 , y otras 3 escalas distintas: y_2, y_3 y y_4). La Figura 1 ilustra el modelo.

Figura 1. Modelo de medida para testar la invarianza de escala



La principal ventaja de los modelos de ecuaciones estructurales reside en que permite el análisis de diferentes grados de invarianza. Se asume que la relación entre la respuesta y la sensación evocada es idéntica (por lo que $\lambda_{1,1} = 1$). Como los datos están normalizados al intervalo [0,1], si existe invarianza de escala, el resto de parámetros lambda deberían de ser también igual a 1. Esta situación es llamada en la terminología de ecuaciones estructurales como modelo de indicadores tau-equivalentes. Si también se asume que el error de medida es diferente de cero $(Var(\delta_{1,1}) > 0)$, un nivel más restrictivo de invarianza de escala ocurriría si la varianza del resto de parámetros theta es igual a $Var(\delta_{1,1})$. Se necesita, primeramente, fijar $Var(\delta_{1,1})$ a un valor específico (ver

Hayduk, 1996). Esta situación es conocida como modelo de indicadores paralelos. Una explicación detallada de estos modelos puede encontrarse en Jöreskog y Sörbom (2001).

El modelo ilustrado en la Figura 1 está identificado, por lo que los investigadores pueden testar diferentes modelos considerando varias restricciones en los parámetros estimados, y así analizar los modelos tau-equivalentes y paralelos. Todos estos modelos pueden ser comparados utilizando el test de diferencias de la chi-cuadrado en una secuencia de modelos anidados (Yuan y Bentler, 2004).

Las limitaciones de esta forma de analizar la invarianza de escala están asociadas a las limitaciones de la metodología de los modelos de ecuaciones estructurales (Tomarken y Waller, 2005). En general, es deseable contar con datos continuos para poder aplicar el método de máxima verosimilitud. Sin embargo, un tamaño de muestra mínimo es necesario para computar la matriz de covarianzas asintótica y para evitar sesgo en el estadístico chi-cuadrado (Jackson, 2003). Cuando los datos no son continuos, son necesarios métodos de libre distribución, y los requerimientos respecto al tamaño muestral son mucho más restrictivos. En este caso, los datos de entrada no provienen de una matriz de covarianzas, sino de una matriz de correlaciones policórica o poliserial, por lo que la interpretación del error de medida como un porcentaje de la varianza de la variable latente se hace más complejo. Este último hecho, dificulta la fijación de parámetros de la escala libre para testar los modelos tau-equivalente y paralelo.

Además, este tipo de metodología requiere que los modelos estén identificados, por lo que se necesita un número determinado de indicadores por cada modelo susceptible de ser testado. Por ejemplo, el modelo mostrado en la Figura 1 no está identificado con sólo dos indicadores; consecuentemente, una simple comparación entre dos escalas no sería posible.

Análisis de entropía: D-test

Recientemente, Martínez y Ruiz (2008) proponen el *D*-test, como herramienta para analizar la invarianza de escala y que supera alguna de las limitaciones de los métodos anteriormente descritos. El *D*-test se fundamenta en la dinámica simbólica y en la entropía simbólica, y compara la diferencia entre los patrones de respuesta que provienen de dos escalas de medida distintas. Su definición se muestra en el siguiente teorema:

Sea P una población de cardinalidad N tal que cada individuo $e \in P$ evalúa un determinado ítem. Cada individuo da su valoración en una escala A y en una escala B. Denominamos a $h(S)$, $h(S,A)$ y $h(S,B)$ la entropía total simbólica de cada valoración, la valoración realizada con la escala A, y la valoración realizada con la escala B, respectivamente. Si la valoración hecha con la escala A no difiere de la valoración hecha con la escala B, entonces:

$$D(k) = 4N[h(S) - h(S,A) - h(S,B) - \ln\left(\frac{1}{2}\right)] \tag{1}$$

se distribuye asintóticamente como χ^2_{k-1}, es decir como una chi-cuadrado con *k-1* grados de libertad, siendo *k*, el número de particiones realizadas al intervalo [0,1]

Sea α un número real con $0 \le \alpha \le 1$. Sea χ^2_α tal que:

$$P(\chi^2_{k-1} > \chi^2_\alpha) = \alpha \tag{2}$$

Entonces la regla de decisión en la aplicación del *D*-test al nivel de confianza $100(1-\alpha)\%$ es:

$$\begin{array}{c} \text{Si } 0 \le D(k) \le \chi^2_\alpha \text{ Aceptar } H_0 \\ \text{de otro modo, Rechazar } H_0 \end{array} \tag{3}$$

siendo H_0 la hipótesis de partida: existe invarianza de escala.

La principal ventaja de este test estriba en que no requiere ninguna asunción sobre la distribución de los datos. Por el contrario, el principal requerimiento es que el tamaño de la muestra sea al menos 10 veces superior al número de particiones realizadas. Como normalmente las particiones deben concordar con el rango de la escala con el menor número de alternativas de respuesta, entonces esta limitación no es muy exigente. No obstante, cuando se trabaja con muestras no muy grandes, puede que el test no tenga potencia suficiente para detectar efectos que perturben la hipótesis de invarianza. Al no reportar intervalos de confianza, este hecho puede convertirse en un problema.

Diferencia entre el área de dos funciones (I)

Se propone a continuación un nuevo y fácilmente comprensible método para estudiar la invarianza de escala, que se basa en la computación de la diferencia de áreas entre dos funciones. Este método permite considerar la imprecisión de las estimaciones a través de intervalos de confianza, y da una medida de tamaño de efecto de la desviación de la invarianza de escala. A continuación se describe su fundamento.

Se parte de la premisa de que si existe invarianza de escala entre dos escalas A y B, siendo A la escala de referencia, entonces se puede encontrar un hiperplano que pase por el origen y que se ajuste perfectamente entre las respuestas de *A* y *B*. Es decir, existe una función lineal que se ajusta sin error a los datos, y que tiene su origen en el punto (0,0). Como los datos están normalizados, la representación en el plano cartesiano es tal y como ilustra la Figura 2, dividiendo la superficie en dos mitades idénticas, por lo que el área por debajo de la curva de ajuste es 0.5.

Formalmente, se plantea un simple modelo de regresión lineal, que puede ser estimado a través de mínimos cuadrados ordinarios si se cumplen las condiciones necesarias (ver Wooldridge, 2006) (4):

$$\begin{aligned} &A = \alpha + \beta B + \varepsilon \\ &f(\bullet) = E(A \mid B) = \hat{\alpha} + \hat{\beta} B = B \end{aligned} \tag{4}$$

siendo A y B los vectores de respuestas para las escalas propuestas, ε el término de error ($\varepsilon = 0$), $\hat{\alpha}$=0 y $\hat{\beta}$=1, ya que debe existir una relación idéntica entre ambas respuestas. Llamaremos a la función ideal: $f(\bullet)$.

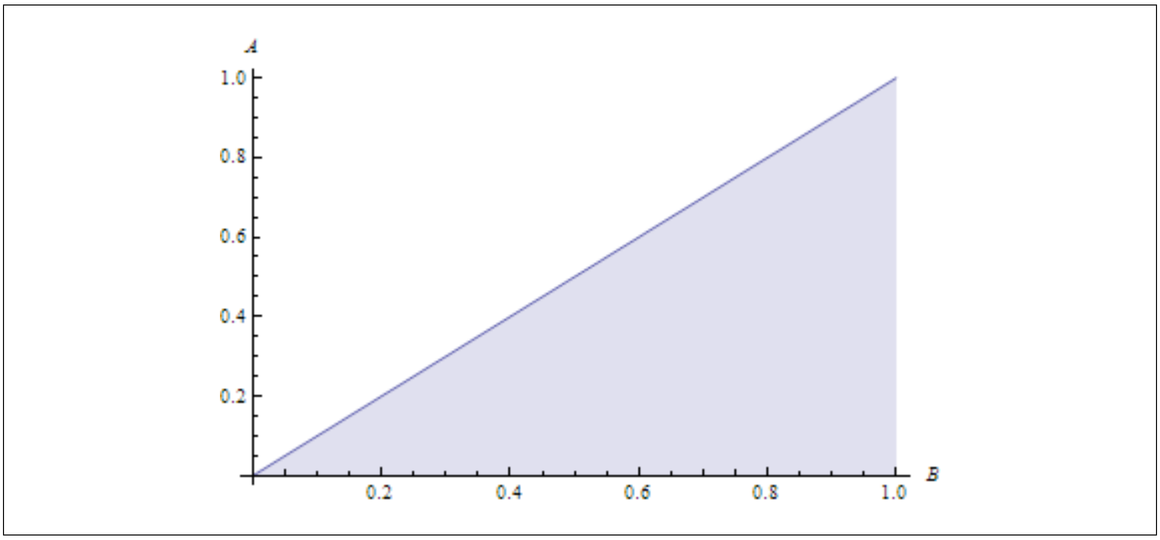


Figura 2. Función invariante entre las escalas A y B

Esta relación ideal de identidad no se dará en la práctica, ya que aunque haya invarianza de escala, la consideración del error de medida y del error por categorización haría que existiera una diferencia entre la función empírica y la función ideal.

La diferencia entre ambas funciones puede establecerse comparando el área que las delimita con el área bajo la curva ideal. De este modo, a medida que exista una mayor desviación de la situación de invarianza, la superficie delimitada entre las funciones se aproximará a la superficie bajo la curva ideal. Así, una función totalmente divergente de la función ideal formará un área=0.5[1], que es precisamente el área bajo la curva ideal, por lo que la máxima desviación de la invarianza ocurrirá cuando el ratio entre ambas áreas sea igual a 1.

Por tanto, se podría graduar la desviación de la situación ideal de invarianza, obteniendo un índice en una escala [0,1] que cuantificara esa desviación. Este índice podría ser interpretado como la magnitud del efecto de interacción de la escala, es decir, como un valor de tamaño de efecto que relativizara esa diferencia. De este modo, se podría evitar los problemas asociados a la interpretación de test estadísticos con niveles de potencia muy dispares, donde tamaños de efecto muy pequeños podrían ser significativos o no, en función del tamaño muestral.

Formalmente, y dada una función empírica $g(\bullet)$:

$$A^*(0,1) = \sum \int_a^b |(f(\bullet) - g(\bullet))|\, dx = \begin{cases} \dfrac{A^*}{A} = I = 0 \quad , \quad \exists \text{ invarianza} \\ 0 < I < 1 \quad , \quad \text{graduación de invarianza} \\ I = 1 \quad , \quad \text{no } \exists \text{ invarianza} \end{cases} \tag{5}$$

siendo A^* el área entre las dos funciones, empírica e ideal, A el área bajo la curva ideal, e *I* el cociente entre ambas. Los límites de integración son *b* y *a,* y representan los puntos de corte de las funciones. Así, cuando una función corte una vez a la función ideal en el punto *z*, se resolverá el cálculo de la siguiente forma: $\int_0^z |(f(\bullet)-g(\bullet))|\,dx+\int_z^1 |(f(\bullet)-g(\bullet))|\,dx$ La Figura 3 ejemplifica gráficamente A^*, dada la función $g(\bullet)=0.165+0.73B$

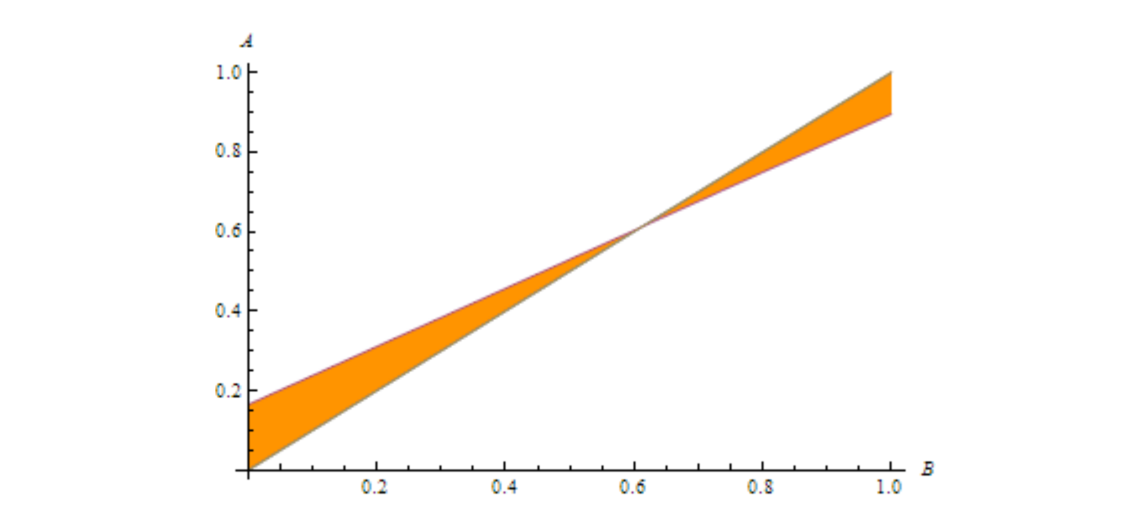


Figura 3. Área entre la función empírica y la función ideal

La flexibilidad de este método permite considerar los intervalos de confianza de la estimación empírica de $\hat{\alpha}$ y $\hat{\beta}$, por lo que se puede trasladar esa cuantificación de la imprecisión de las estimaciones al valor de *I*. Es decir, se puede reportar un valor puntual de *I* y un intervalo de confianza $(I_{\inf}, I_{\sup})$. Para ello, hay que calcular los valores $(A^*_{\inf}, A^*_{\sup})$. La Figura 4 ejemplifica gráficamente esos valores para las funciones: $g(\bullet)_{\inf}=0.08+0.62B$ y $g(\bullet)_{\sup}=0.25+0.84B$.

Una cuestión importante a la hora de interpretar esos intervalos es el hecho de que $A^*_{\inf}$ y $A^*_{\sup}$ pueden ser mayores que A^*, ya que se pueden alejar en mayor medida de la recta ideal. Es decir *I* puede no estar incluida en $(I_{\inf}, I_{\sup})$. Ello puede ocurrir en función del grado en que $\hat{\beta}$ se aleje de 1, o lo que es lo mismo, cuando $g(\bullet)$ se vaya alejando de la forma paralela y acercando a la ortogonal con respecto a $f(\bullet)$. Este particular hecho lleva a considerar esos intervalos como *intervalos de tolerancia* (IT), para no confundir con la habitual interpretación estadística de aquéllos, donde siempre el intervalo de confianza (IC) contiene al parámetro.

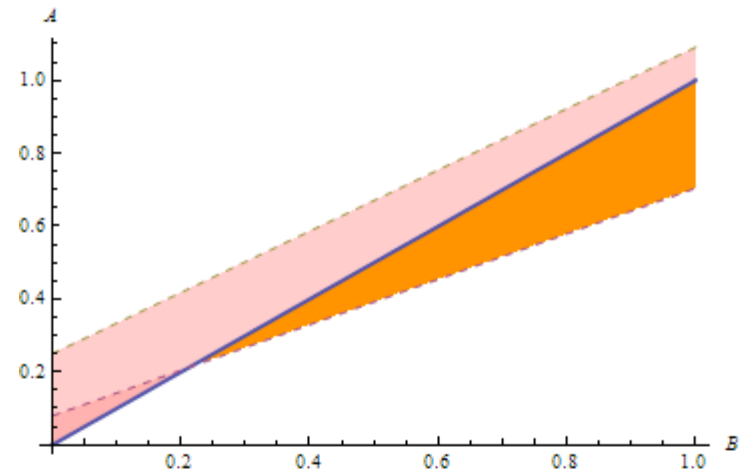


Figura 4. Áreas considerando los intervalos de tolerancia

El error de medida de la variable independiente puede tenerse en cuenta asumiendo un cierto valor de fiabilidad r_B para esa medición. Ree y Carreta (2006) ofrecen guías para realizar ese cálculo, consistente en corregir los parámetros estimados $\hat{\alpha}$ y $\hat{\beta}$, por $\hat{\alpha}_c$ y $\hat{\beta}_c$. De este modo, se relaja una de las asunciones comúnmente especificadas en este tipo de modelos.

$$\begin{aligned}\hat{\beta}_c &= \frac{\hat{\beta}}{r_B} \\ \hat{\alpha}_c &= \overline{A} - \hat{\beta}_c\overline{B}\end{aligned} \qquad (6)$$

siendo $\overline{A}$ y $\overline{B}$ los valores medios de las respuestas para ambas escalas.

El método de los mínimos cuadrados ordinarios es robusto frente a desviaciones de normalidad de los errores, por lo que realmente la asunción más importante sobre la que descansa esta estimación es la forma de la relación entre ambas variables, que se asume lineal, además de la covariación nula entre el predictor y el error. La homocedasticidad de los errores es también otra asunción que puede relajarse, si no se cumple, estimando por mínimos cuadrados ponderados.

Se puede establecer, asimismo, ciertas recomendaciones heurísticas sobre cómo valorar esa graduación de invarianza. Así, podría establecerse un intervalo para I donde se admitiera la existencia de invarianza. Para ello, se tomaría como criterio la premisa de que dos escalas invariantes que miden el mismo fenómeno lo seguirán siendo si existe un pequeño error de medida en una de ellas. Si se corrigen los parámetros $\hat{\alpha}$ y $\hat{\beta}$ por un coeficiente de fiabilidad del 85%, y asumiendo que las medias de ambas escalas coinciden, entonces se llega a la siguiente ecuación: $g(\bullet) = -0.176 + 1.176B$.

Comparando esa función con la función ideal, se obtiene un valor de $A^* = 0.088$, por lo que $I = 0.175$. Por tanto, en este rango de valores: $0 \le I \le 0.175$, se podría considerar que existe invarianza, es decir, si I o el intervalo de tolerancia de I incluye algún valor de ese rango, existirá invarianza de escala.

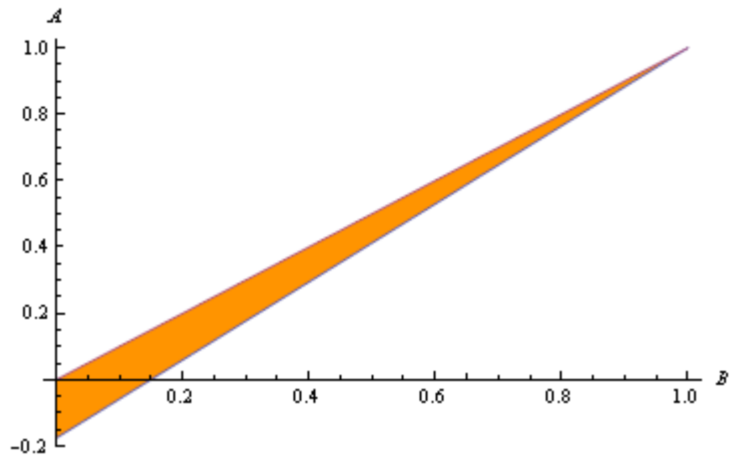


Figura 5. Desviación permisible de la invarianza

La Tabla 1 resume la información más relevante de cada método comentado, sus ventajas e inconvenientes.

Tabla 1. Métodos para analizar la invarianza de escala

	Test estadístico	Tamaño de efecto	Principales ventajas	Principales inconvenientes
Comparación directa de medias y varianzas	Sí	Sí	- Simplicidad	- Requiere que la distribución de respuestas sea uniforme o perfectamente simétrica para balancear errores - Datos agregados - Puede admitirse invarianza aunque el patrón de respuestas sea claramente invariante
Modelos de ecuaciones estructurales	Sí	No	- Test del modelo - Dos niveles de invarianza - Consideración del error de medida	- Tamaño de muestra mínimo - Requiere constrir un modelo de al menos 4 indicadores - Difícil la detección de un indicador problemático cuando hay errores de especificación - Datos continuos y normales multivariables para estimar por máxima verosimilitud
Análisis de entropía (*D*-test)	Sí	No se especifica	- Sin asunciones sobre la distribución y naturaleza de los datos	- Tamaño de muestra superior a 10 veces el número de particiones realizadas - No consideración error de medida
Diferencia entre el área de dos funciones (*I*)	No	Sí	- Consideración del error de medida - Graduación de la invarianza	- No existe un criterio estadístico para la graduación de invarianza, sino heurístico. - El cálculo integral puede resultar tedioso

El estudio sujeto a re-análisis

El estudio de Martínez y Martínez (2008b) consta de dos muestras aleatorias (la segunda de ellas de replicación) recogidas en los dos centros deportivos más importantes de la ciudad de Cartagena.

La primera de esas muestras la formaron 116 consumidores, constituyendo alrededor del 6% de usuarios de esos servicios. El factor de imprecisión sobre la escala de medida- *FIEM*- (Martínez y Martínez, 2008a) 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.

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. 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.

La recogida de datos se realizó durante la primavera de 2008, de forma secuencial en cada centro deportivo. Ambos centros ofrecen una gran diversidad de actividades físico-deportivas. El primero de ellos oferta una gama de 20 actividades monitorizadas (musculación, aeróbic, danza, gimnasia rítmica, etc.) además del servicio de alquiler de pistas polideportivas. El segundo, por su parte, tiene un portafolio de servicios más reducido porque se enfoca principalmente en la oferta de actividades acuáticas, aunque también existen otras como musculación, aeróbic, defensa personal, etc., además de la posibilidad de alquiler de pistas. Esta segunda instalación es de gestión compartida con una empresa privada, por lo que Martínez y Martínez (2008b) asumieron que podría existir heterogeneidad a priori, y por tanto, diferenciación en la situación del test.

Dos estudiantes de marketing previamente entrenados y formados para la realización de encuestas fueron los encargados de la recogida de datos. Para ello, debían entrevistar individualmente a usuarios en la entrada de ambos centros, seleccionando esos individuos de forma aleatoria. En realidad, esta es una situación de pseudo-aleatoriedad, ya que la selección de individuos depende de cuándo el encuestador haga las entrevistas, por lo que se especificó que los encuestadores debían de realizar esa recogida de datos en varias etapas, representando cada etapa una fecha diferente (día y hora). De esta manera, se intentaba minimizar la presencia de sesgo de selección de muestreo.

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

Tras dos preguntas que no versaban sobre calidad percibida, se les volvía a preguntar sobre su percepción de calidad del servicio utilizando (y en el orden siguiente) las escalas categóricas mencionadas anteriormente: diferencial semántico de -3 a +3, Likert de 1 a 5 y de 1 a 7. Martínez y Martínez (2008b) justifican el porqué de la elección de esas escalas. 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 reducir la posibilidad de que existiera dependencia entre las respuestas numéricas, lo que representaría un problema a nivel de análisis estadístico.

Las preguntas sociodemográficas completaban el cuestionario final, que era muy sencillo y rápido de contestar.

Resultados

Los datos del estudio de Martínez y Martínez (2008b) fueron re-analizados utilizando los cuatro métodos descritos. Los resultados para las dos muestras (inicial y replicación) se indican a continuación.

Comparación directa de medias y varianzas

Esta primera visión de los datos indica que existe bastante homogeneidad en los estadísticos (Tabla 2). No obstante, la escala Likert de 1 a 5 se atisba que tiene un comportamiento ligeramente más alejado del resto. Por tanto, no existe una heterogeneidad evidente en las medias y varianzas, aunque sí es cierto que esta segunda muestra tiene una distribución mucho menos apuntada que la primera, y se acerca más a la normalidad multivariante.

Tabla 2. Comparación directa entre medias y varianzas

Muestra inicial*	Media (IC 95%)	Varianza (IC 95%)	Invarianza
Escala libre	0.710 (0.669 ; 0.751)	0.050 (0.039 ; 0.066)	
Diferencial semántico	0.743 (0.699 ; 0.787)	0.057 (0.044 ; 0.075)	Sí
Likert de 1 a 5	0.643 (0.598 ; 0.688)	0.062 (0.048 ; 0.081)	Sí
Likert de 1 a 7	0.694 (0.650 ; 0.738)	0.057 (0.044 ; 0.075)	Sí
Replicación**			
Escala libre	0.677 (0.639 ; 0.715)	0.036 (0.027 ; 0.048)	
Diferencial semántico	0.781 (0.701 ; 0.781)	0.040 (0.030 ; 0.054)	Sí
Likert de 1 a 5	0.608 (0.563 ; 0.653)	0.052 (0.040 ; 0.070)	Sí
Likert de 1 a 7	0.702 (0.618 ; 0.702)	0.044 (0.034 ; 0.060)	Sí

* Kurtosis multivariante relativa: 1.77
** Kurtosis multivariante relativa: 1.21

Modelos de ecuaciones estructurales

La Tabla 3 muestra los resultados de la secuencia de los modelos anidados. Se han reproducido los análisis originales de Martínez y Martínez (2008b), añadiendo además varias asunciones sobre fiabilidad. Como puede contemplarse, el modelo base y el de ítems tau-equivalente se ajustan, pero no el modelo de indicadores paralelos ($M_{3:85}$). Sin embargo, $M_{2:85}$ tiene un ajuste comparativamente peor que $M_{1:85}$, ya que $\Delta SB\chi 2(gl)$ es significativa (p<0.05). No obstante, $M_{2:85}$ es un modelo estadísticamente aceptable, aunque en la frontera de la significación.

Sin embargo, la sensibilidad del modelo a la fiabilidad de la escala es importante, ya que los modelos $M_{1:90}$ y $M_{1:100}$ no se ajustan, por lo que ni siquiera se cumple el principio de independencia condicional, es decir, la configuración del modelo no es correcta. Como estos modelos base no está correctamente especificado, no tienen sentido las comparaciones relativas entre modelos (Yuan y Bentler, 2004). En la muestra de replicación, ninguno de los modelos base se ajusta, por lo que es indicativo de mala especificación, es decir, existe evidencia de que no sólo al menos algún ítem no es invariante, sino que puede haber problemas de dependencia entre ítems.

Tabla 3. Modelos estadísticos con LISREL

Fiabilidad escala libre	Modelo	Restricciones	*SB*χ^2 *(gl)*	*p* valor	*Δ SB*χ^2*(gl)*	*p* valor
85%	$M_{1:85}$	Modelo base	4.34 (3)	0.23		
90 %	$M_{1:90}$	Modelo base	8.00 (9)	0.04		
100%	$M_{1:100}$	Modelo base	35.21 (3)	0.00		
85%	$M_{2:85}$	Modelo de items tau-equivalentes	12.55 (6)	0.05	$M_{2:85}$- $M_{1:85}$=21.43 (3)	0.00
85%	$M_{3:85}$	Modelo de items paralelos	19.75 (9)	0.02		
	Replicación					
85%	$Mr_{1:85}$	Modelo base	19.41 (3)	0.00		
90 %	$Mr_{1:85}$	Modelo base	32.18 (3)	0.00		
100%	$Mr_{1:85}$	Modelo base	77.68 (3)	0.00		
85%	$Mr_{2:85}$	Modelo de items tau-equivalentes	36.17(6)	0.00		
85%	$Mr_{3:85}$	Modelo de items paralelos	53.97 (9)	0.00		

Análisis de entropía (D-test)

En este caso, los resultados muestran cómo las escalas de 7 opciones de respuesta, tanto Likert como diferencial semántico, son invariantes (Tabla 4).

La replicación del *D*-test muestra resultados consistentes con los derivados de la primera muestra. De nuevo hay evidencias que favorecen la hipótesis de que existe invarianza de escala para las opciones de medición con 7 categorías de respuesta.

Tabla 4. Aplicación del *D*-test

Comparación[a]	*D*-test	Subintervalos (k)	Valores de corte de la Chi-cuadrado (95%)	Invarianza
A-B	0.94	7	12.59	Sí
A-C	14.88	5	9.48	No
A-D	6.96	7	12.59	Sí
Replicación				
A-B	3.37	7	12.59	Sí
A-C	12.61	5	9.48	No
A-D	9.61	7	12.59	Sí

[a] Escala libre (A); Diferencial semántico de -3 a +3 (B); Likert de 1 a 5 (C); Likert de 1 a 7 (D)

Diferencia entre el área de dos funciones (I)

Los resultados muestran cómo de nuevo la escala de 1 a 5 tiene un comportamiento no invariante (Tabla 5). Bien es cierto que para niveles de fiabilidad del 85 y 90% el valor de *I* está dentro del límite en la muestra inicial, pero no así en la de replicación.

Tabla 5. Aplicación de la diferencia entre el área de dos funciones (*I*)

Fiabilidad	Comparación[a]	$\hat{\alpha}$ (95% IC)	$\hat{\beta}$ (95% IC)	A^* (95% IT)	I (95% IT)	Invarianza
r_B = 85%	A-B	0.069 (-0.014 ; 0.152)	0.861 (0.753 ; 0.970)	0.034 0.137 ; 0.137	0.068 (0.274 ; 0.274)	Sí
r_C = 85%	A-C	0.163 (0.097 ; 0.230)	0.848 (0.750 ; 0.947)	0.087 (0.064 ;0.203)	0.174 (0.128 ; 0.406)	Sí
r_D = 85%	A-D	0.046 (-0.004 ; 0.096)	0.956 (0.886 ; 1.026)	0.024 (0.061 ; 0.109)	0.048 (0.122 ; 0.218)	Sí
r_B = 90%	A-B	0.104 (0.020 ; 0.189)	0.814 (0.704 ; 0.923)	0.047 (0.129 ; 0.135)	0.094 (0.258 ; 0.270)	Sí
r_C = 90%	A-C	0.194 (0.125 ; 0.262)	0.801 (0.701 ; 0.902)	0.094 (0.077 ; 0.212)	0.188 (0.154;0.424)	Sí
r_D = 90%	A-D	0.083 (0.026 ; 0.140)	0.903 (0.824 ; 0.982)	0.036 (0.066 ; 0.131)	0.072 (0.132 ; 0.262)	Sí
r_B = 100%	A-B	0.165 (0.080 ; 0.250)	0.732 (0.624 ; 0.841)	0.070 (0.124 ; 0.170)	0.140 (0.248 ; 0.340)	Sí
r_C = 100%	A-C	0.245 (0.176 ; 0.315)	0.721 (0.620 ; 0.822)	0.109 (0.094 ; 0.260)	0.218 (0.188 ; 0.520)	No
r_D = 100%	A-D	0.145 (0.082 ; 0.209)	0.813 (0.726 ; 0.899)	0.060 (0.078 ; 0.158)	0.120 (0.156 ; 0.316)	Sí
	Replicación					
r_B = 85%	A-B	0.060 (-0.038 ; 0.160)	0.832 (0.701 ; 0.963)	0.014 (0.188 ; 0.141)	0.028 (0.376 ; 0.282)	Sí
r_C = 85%	A-C	0.232 (0.157 ; 0.306)	0.731 (0.615 ; 0.848)	0.102 (0.100 ; 0.230)	0.204 (0.200 ; 0.460)	No
r_D = 85%	A-D	0.131 (0.051 ; 0.211)	0.827 (0.710 ; 0.943)	0.055 (0.102 ; 0.182)	0.110 (0.204; 0.364)	Sí
r_B = 90%	A-B	0.095 (-0.004 ; 0.194)	0.786 (0.655 ; 0.916)	0.054 (0.176 ; 0.152)	0.108 (0.352 ;0.304)	Sí
r_C = 90%	A-C	0.256 (0.182 ; 0.331)	0.691 (0.575 ; 0.807)	0.110 (0.108 ; 0.231)	0.220 (0.216 ; 0.462)	No
r_D = 90%	A-D	0.162 (0.081 ; 0.242)	0.781 (0.663 ; 0.898)	0.067 (0.106 ; 0.191)	0.134 (0.212 ; 0.382)	Sí
r_B = 100%	A-B	0.153 (0.056 ; 0.250)	0.707 (0.580 ; 0.834)	0.073 (0.163 ; 0.167)	0.146 (0.326 ; 0.334)	Sí
r_C = 100%	A-C	0.298 (0.225 ; 0.371)	0.622 (0.509 ; 0.734)	0.125 (0.121 ; 0.238)	0.250 (0.242 ; 0.476)	No
r_D = 100%	A-D	0.213 (0.133 ; 0.293)	0.703 (0.587 ; 0.819)	0.087 (0.115 ; 0.202)	0.174 (0.230 ; 0.404)	Sí

[a] Escala libre (A); Diferencial semántico de -3 a +3 (B); Likert de 1 a 5 (C); Likert de 1 a 7 (D).
Valor de corte: I = 0.175

Discusión

Esta investigación presenta un nuevo método para el estudio de la invarianza basado en el cálculo integral, y que complementa otros métodos existentes, lo que contribuye a realizar un análisis ecléctico y multifocal de un problema importante en la investigación de marketing.

Dado que la investigación en gestión y marketing del deporte está cobrando una mayor relevancia en el ámbito académico, la preocupación por implementar mediciones válidas y útiles de los fenómenos de interés es primordial. El estudio de Martínez y Martínez (2008b) es una aportación novedosa en este campo, y en concreto sobre la investigación sobre calidad percibida, abogando por el uso de la perspectiva en primera persona, es decir, el consumidor reporta su valoración en una escala libre, que es la que maximiza la utilidad de las respuestas.

Como muestran estos autores, los consumidores de servicios deportivos prefieren dar sus respuestas en escalas de 0 a 10 o de 1 a 10, en lugar de en las que habitualmente se indican en los cuestionarios, encontrando además, que los formatos de respuesta: Likert de 1 a 5, Likert de 1 a 7, y diferencial semántico de -3 a +3 no son invariantes, por lo que producen una distorsión no aceptable sobre los resultados. De esos resultados pueden derivarse conclusiones como que las investigaciones sobre calidad percibida (o incluso otras actitudes similares, como la satisfacción) que se hayan realizado utilizando esas escalas tienen un sesgo a tener en cuenta. Sin embargo, la aplicación del enfoque multimétodo realizado en este trabajo muestra que otra conclusión menos lapidaria es también posible.

Tanto los resultados del *D*-test, como los derivados del cálculo integral, indican que la escala Likert de 1 a 7 y la diferencial semántico pueden considerarse invariantes. La escala Likert de 1 a 5, por el contrario, sigue teniendo un comportamiento anómalo, es decir, la restricción por categorización afecta ostensiblemente los resultados.

Las implicaciones para la gestión deportiva que se derivan de estos nuevos resultados son importantes, ya que, por un lado, vuelven a incidir en el peligro de las escalas Likert de 1 a 5 (las más utilizadas), y por otro lado indican que las escalas de 7 opciones de respuesta no producen una distorsión relevante. Así, destacadas investigaciones en esta disciplina pueden estar sujetas a ese sesgo, ya que utilizaron una escala no invariante (ej. Hernández, 2001; Morales, Hernández y Blanco, 2005), mientras que otras no (ej. Ko y Pastore, 2005). Los primeros autores implementaron mediciones con escalas de cinco categorías, mientras que los segundos con escalas de siete. No obstante, las recomendaciones de Martínez y Martínez (2008b) acerca de utilizar la perspectiva en primera persona siguen vigentes, al menos hasta que posteriores estudios muestren de forma más consistente si existe invarianza en esos formatos de respuesta, y pueda realizarse un metanálisis.

El método propuesto presenta ventajas con respecto a los modelos de ecuaciones estructurales y al análisis de entropía, y tiene gran flexibilidad. Sin embargo, su carácter heurístico hace que el valor de corte definido para I pueda criticarse por parecer arbitrario. Se ha justificado, no obstante, ese valor a través de la relación con la fiabilidad de la escala, lo que le da robustez. El valor de fiabilidad de 85% sí que puede estar sujeto a variaciones, en función del contexto de investigación, lo que haría cambiar el valor de corte para I.

En cualquier caso, se debe exigir valores de fiabilidad en torno a ese número (ver Hayduk, 1996), por lo que el valor propuesto parece ser razonable. La consistencia de resultados con el *D*-test, favorece la aceptación de éste nuevo método, añadiendo las ventajas que éste conlleva, como la interpretación geométrica y la consideración del error de medida y los intervalos de tolerancia.

Lo tedioso del cálculo integral puede complicar la diseminación de este nuevo método, aunque la ayuda de programas informáticos, como por ejemplo *Mathematica* (que es el que se ha utilizado en esta investigación), facilitará al investigador de forma ostensible esos cálculos.

Finalmente, es importante incidir en la idea de la necesidad de la triangulación estadística, que como ha podido comprobarse en este estudio, ayuda a tener una visión más global sobre los resultados, y evita derivar conclusiones que pueden ser parcialmente erróneas. Excepto del primer método utilizado (que tiene muchas debilidades), de los otros tres se puede obtener información relevante y complementaria. Por ejemplo, en la aplicación de las ecuaciones estructurales, el hecho de que en la muestra de replicación no se ajuste el modelo base es indicativo de la no existencia de independencia condicional, por lo que cabría la posibilidad de que un ítem estuviera también afectado por otra variable latente (en este caso podría ser por el propio efecto método de la restricción de escala). Ello podría testarse empíricamente si se tuvieran más grados de libertad, y plantear un modelo invariante en las escalas de 7 opciones y no invariante en la de 5, el cual, muy probablemente, se ajustaría de forma aceptable.

Conclusiones

La invarianza de escala es una condición necesaria para que los instrumentos de medida utilizados no produzcan distorsión en la representación del fenómeno de interés. Es por ello que se hace imprescindible su análisis cuando se utilizan escalas de medida que no coinciden con las escalas preferidas por los individuos para dar sus respuestas. El nuevo método propuesto, basado en la delimitación del área entre dos funciones, es una forma heurística y flexible para realizar ese análisis, y que complementa los otros métodos existentes.

Como esta investigación ha demostrado, medir la calidad percibida con escalas Likert de 1 a 7 y diferencial semántico de -3 a +3 no produce sesgo en los resultados, es decir, existe invarianza de escala. Por el contrario, la escala Likert de 1 a 5 sí interactúa con las respuestas de los individuos, por lo que contamina le medición de su percepción de calidad. Aunque más investigaciones son necesarias para establecer si estos resultados pueden considerarse como norma, se puede afirmar que los individuos prefieren dar sus evaluaciones de calidad sobre una referencia de 10 puntos, y que si el investigador propone un instrumento de medida que obligue al encuestado a categorizar de forma importante su respuesta (utilizar una escala de 5 puntos), entonces se produce una distorsión relevante de ésta.

Limitaciones

Además de las limitaciones del método propuesto comentadas anteriormente (no existe un criterio estadístico, sino heurístico, para la graduación de la invarianza, y que el cálculo integral puede resultar tedioso), las conclusiones de esta investigación están condicionadas principalmente por la metodología utilizada en el estudio de Martínez y Martínez (2008b).

En su artículo original, Martínez y Martínez (2008b), ya comentan que la principal limitación reside en la posibilidad de que existan modelos competitivos de dependencia entre indicadores. Aunque trataron de minimizar ese posible riesgo en el diseño del cuestionario, estos autores admiten 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 utilizando ecuaciones estructurales sería necesario incrementar el número de variables observables utilizadas, con el fin de poder identificar el modelo. Si esa dependencia realmente existiera, los resultados quedarían sesgados. No obstante, concordamos con Martínez y Martínez (2008b) en que la forma bien explicitada de realizar la encuesta a cada individuo, minimiza ostensiblemente ese riesgo.

Al realizar la comparación por pares de escalas que utilizan el *D*-test y el método basado en el cálculo integral, la existencia de un efecto de orden podría también ser una fuente de sesgo. Ese efecto de orden se refiere a la distorsión producida en las respuestas por ordenar las preguntas de una determinada manera, y se ha mostrado relevante en la investigación sobre calidad del servicio (DeMoranville y Bienstock, 2003). Es por ello que hubiera sido más recomendable seguir una secuencia variante en las preguntas relativas a la percepción de calidad sobre las tres escalas bajo análisis, es decir, no seguir una secuencia fija en las preguntas una vez que el individuo hubiera realizado su evaluación sobre la escala libre. De este modo, se podría haber aleatorizado la secuencia de las tres preguntas posteriores, para intentar evitar ese posible efecto de orden.

Finalmente, Martínez y Martínez (2008b) defienden la visión metodológica de la medición de constructos a través de un solo ítem, en consonancia con las perspectivas de un emergente grupo de investigadores (ej. Hayduk, 1996; Rossiter, 2002; Hayduk, Pazderka-Robinson, Cummings, Boadu, Verbeek, y Perks, 2007). Es por ello que miden la calidad percibida utilizando una única pregunta de valoración global. Ese tipo de preguntas de valoración global han sido utilizadas por otros autores, como Teas (1993). Ciertamente, esta es una cuestión muy debatida actualmente en la literatura especializada, lo que puede ser una fuente de crítica desde el punto de vista metodológico por algunos investigadores. Sin embargo, es una postura bien sustentada y fundamentada, por lo que desde esta investigación también se la defiende y apoya.

Notas

[1] Pueden existir funciones que produzcan una diferencia de área con respecto a la función ideal mayor que 0.5, pero sus forma se alejaría muchísimo del comportamiento esperado, por ejemplo, una recta paralela a la ideal con un valor de constante muy alejado del origen. Encontrar funciones de ese tipo sería indicativo de que existe un grave problema en la recogida de datos.

Referencias Bibliográficas

Cohen, P.; Cohen, J.; Aiken, L. & West, S. (1999). The problem of units and the circumstance for POMP. *Multivariate Behavioral Research, 34* (3), 315-346.

DeMoranville C. W. & Bienstock, C. C. (2003). Question order effects in measuring service quality. *International Journal of Research in Marketing, 20*, 217–231

Hayduk, L. A. (1996). *LISREL Issues, Debates and Strategies*. Baltimore, MA: Johns Hopkins University Press.

Hayduk, L. A.; Pazderka-Robinson, H.; Cummings, G. G.; Boadu, K.; Verbeek, E. L. & Perks, T. A. (2007). The weird world, and equally weird measurement models: reactive indicators and the validity revolution. *Structural Equation Modeling, 14* (2), 280-310.

Hernández, A. (2001). Un cuestionario para evaluar la calidad en programas de actividad física. *Revista de Psicología del Deporte, 10* (2), 179-196.

Jackson, D. L. (2003). Revisiting sample size and number of parameter estimates: Some support for the N:q hypothesis. *Structural Equation Modeling, 10* (1), 128-141.

Jöreskog, K. G. & Sörbom, D. (2001). *LISREL 8: User's referente guide*. Scientific Software International, Inc.

Ko, Y.J. & Pastore, D. L. (2005). A hierarchical model of service quality for the recreational sport industry*. Sport Marketing Quarterly, 14* (2), 84–97.

Martínez, J. A. y Martínez, L. (2008a). 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, J. A. y Martínez, L. (2008b). 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, 8* (31), 244-255.

Martínez, J. A. & Ruiz, M. (2008). Improving measurement in management and marketing research; a new test for analysing scale invariance using symbolic dynamics and symbolic entropy. *Management Decision*. En revisión.

Morales, V.; Hernández, A. y Blanco, A. (2005). Evaluación de la calidad en los programas de actividad física. *Psicothema, 17* (2), 292-298.

Ree, M. J. & Carreta, T. H. (2006). The role of measurement error in familiar statistics. *Organizational Research Methods, 9* (1), 99-112.

Rossiter, J. R. (2002). The COARSE procedure for scale development in marketing. *International Journal of Research in Marketing, 19* (4), 305-335.

Teas, R. K. (1993). Expectations, performance evaluation and consumer's perceptions of quality. *Journal of Marketing, 57* (Octubre), 18-34

Tomarken, A. J. & Waller, N. G. (2005). Structural equation modeling: Strengths, limitations, and misconceptions. *Annual Review of Clinical Psychology, 1*, 31-65.

Tsitskari. E.; Tsiotras, D. & Tsiotras, G. (2006). Measuring service quality in sport services. *Total Quality Management & Business Excellence, 17* (5), 623-631

Tukey, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley

Wooldridge, J. M. (2006). *Introducción a la Econometría. Un enfoque moderno*. Segunda edición. Thomson.

Yuan, K. H. & 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* (5), 737–757.