

Aplicación de la teoría de la generalizabilidad en la evaluación de servicios en gestión deportiva

Application of the generalizability theory in the services evaluation in sports management

Pablo Gálvez-Ruiz^{1*}, Moisés Grimaldi-Puyana², Jesús Fernández-Gavira² y Antonio Jesús Sánchez-Oliver^{2,3}

1 Departamento de Educación en Internet. Facultad de Educación en Internet. Universidad Internacional de La Rioja (España).

2 Departamento de Educación Física y Deporte. Facultad de Educación. Universidad de Sevilla (España).

3 Departamento de Deporte e Informática. Facultad de Deporte. Universidad Pablo de Olavide (España).

Resumen: Esta investigación aborda a partir de la Teoría de la Generalizabilidad (TG) la optimización de recursos en la gestión deportiva. Se utilizó una estructura de diseño compuesta por las facetas usuarios (U), actividades (A), escalas (E) e ítems (I), siguiendo un diseño no ortogonal, parcialmente anidado y multifaceta dentro del Modelo General Lineal (GLM). Los 537 participantes (32.11 ± 11.11 años) cumplieron el Cuestionario de Evaluación de la Calidad Percibida en Servicios Deportivos obteniendo resultados excelentes de fiabilidad ($\xi\rho^2(\delta) > 0.98$) y generalizabilidad ($\xi\rho^2(\Delta) > 0.98$). La optimización mostró que con una reducción del número de participantes a la mitad y con un menor número de actividades evaluadas, obtenemos un alto grado de generalización al mismo tiempo que se reduce el número total de observaciones de cara a futuros registros, lo que facilitará la evaluación del servicio deportivo de manera más eficiente.

Palabras clave: gestión deportiva, análisis de generalizabilidad, evaluación

de la calidad, calidad percibida

Abstract: This paper addresses from the use of the Generalizability Theory (TG) the optimization of resources in sports management. We used a design structure composed of the facets users (U), activities (A), scales (E) and items (I), following not orthogonal design, partially nested and multifaceted, within the General Linear Model. 537 participants (32.11 ± 11.11 years) completed the Questionnaire of Evaluation of the Perceived Quality in Sports Services obtaining excellent results both indexes of reliability ($\xi\rho^2(\delta) > 0.98$) and generalizability ($\xi\rho^2(\Delta) > 0.98$). The optimization showed that half of participants and fewer activities, we obtain a high degree of generalization and at the same time, reduced the total number of observations, being important for future studies which will facilitate the evaluation of sport service more efficiently.

Keywords: sports management, generalizability analysis, evaluation of the quality, perceived quality.

Introducción

La investigación en torno a la calidad percibida en servicios deportivos ha ofrecido en los últimos años distintas aportaciones de instrumentos de evaluación con el objetivo de mejorar los sistemas de calidad, el servicio ofrecido, y con ello, la satisfacción de los clientes (e.g. Barros y Gonçalves, 2009; Theodorakis, Howat, Ko y Avourdiadou, 2014), superando definitivamente los dos instrumentos más empleados en las últimas tres décadas: SERVQUAL (Parasuraman, Zeithaml y Berry, 1985) y SERVPERF (Cronin y Taylor, 1992). Si bien ambos han sido ampliamente probados, adaptados y replicados en varios entornos y contextos, sus dimensiones no son generalizables y difieren en función del servicio o incluso del entorno geográfico (Batista y Coenders, 2012).

En el caso de los servicios deportivos, el desarrollo de nuevos modelos de negocio dentro de la industria del fitness inciden directamente en la tipología de oferta y evolución de la gestión, lo que conlleva cambios para adaptarse a las nuevas

situaciones. De hecho, el sector de los servicios deportivos difiere en gran medida de otras industrias de servicios puesto que los clientes tienen una mayor frecuencia de encuentros de servicio, y en el caso de las actividades dirigidas, son incluso co-productores del servicio al estar físicamente presentes y ejercer activamente para que el proceso de prestación sea exitoso (Bitner, Faranda, Hubbert y Zeithaml, 1997).

El desarrollo experimentado por este sector está asociado a un crecimiento sostenido tanto a nivel de número de clientes como de centros en todo el mundo. Se estima que los miembros de estas organizaciones actualmente incluyen a 50 millones de europeos (Gocłowska y Piatkowska, 2017), y concretamente en España, solamente la industria del fitness ocupa el quinto mercado más importante y presenta un buen índice de penetración de la industria (EuropeActive y Deloitte, 2017) aunque alejado de las primeras posiciones. Por ello, la investigación en torno al comportamiento del cliente puede contribuir activamente a la calidad de los servicios (Yi, Gong y Lee, 2013) al tratarse de una de las variables más analizadas junto con la satisfacción y el valor percibido, ya que se considera como el paso anterior a la fidelidad (Baker y Crompton, 2000), y por tanto el camino para la obtención

Dirección para correspondencia [Correspondence address]: Pablo Gálvez-Ruiz. Departamento de Educación en Internet. Facultad de Educación en Internet. Universidad Internacional de La Rioja (España).
E-mail: pablo.galvez@unir.net

de altos índices de productividad y mayores márgenes de beneficio. Además, favorecer la fidelidad del cliente en servicios deportivos estaría en la línea de la promoción de la adherencia a la actividad física que redundaría en menores problemas de salud (Jiménez, Moreno, Leyton y Claver, 2015). Es por este motivo por el que la calidad de servicio ha adquirido una alta importancia para los gerentes de los centros deportivos tanto públicos como privados con el objetivo de detectar qué opinan los clientes y poder establecer acciones correctoras (García-Fernández, Gálvez-Ruiz, Vélez-Colón, Ortega-Fernández y Fernández-Gavira, 2018).

La medición juega un papel importante en la investigación en ciencias del comportamiento (Blanco-Villaseñor, 1991). En el caso de la evaluación de servicios deportivos, se pretende relacionar las percepciones de los usuarios sobre determinados criterios o atributos del servicio en base a las respuestas de una serie de ítems, que sí son observables y por tanto podemos inferir a partir de las respuestas. En este sentido, para medir la relación entre los indicadores empíricos (respuestas) y los conceptos teóricos subyacentes necesitamos conocer la validez, fiabilidad o generalizabilidad, con el objetivo de comprobar las fuentes de error existentes en este proceso.

El uso de herramientas de evaluación, concretamente cuestionarios, debe contar con rigor metodológico para la correcta interpretación de los resultados. Así, entre las propiedades estadísticas que deben sustentar su uso se encuentra la confiabilidad como una de las evidencias de validez relacionada con la estructura interna del instrumento (Espinosa-Vázquez, Martínez-González, Sánchez-Mendiola y Leenen, 2017).

Por ello, ante la especificidad del instrumento la teoría de la generalizabilidad permite medir la confiabilidad de una prueba por medio de la cuantificación de la importancia de cada una de sus fuentes de variabilidad (Zúñiga-Brenes y Montero-Rojas, 2007) a través de las diferentes facetas y las relaciones entre ellas (Cardinet, Johnson y Pini, 2010). De esta manera, se puede establecer la fiabilidad del instrumento así como separar la variabilidad real de la variabilidad del error (Blanco-Villaseñor, 2001; Wickel y Welk, 2010), conseguir el nivel de precisión esperado en el ámbito de medida previsto así como la obtención de los resultados deseados con los mínimos requisitos de esfuerzo y capacidad (ISO 10075-3, 2004).

En el análisis de generalizabilidad se distinguen diversas etapas (Cronbach, Gleser, Nanda y Rajaratnam, 1972): a) el estudio de generalizabilidad (estudio G) está diseñado para aislar y estimar tantas facetas del error de medición como sea razonable y económicamente factible (Webb y Shavelson, 2005), y b) el estudio de decisión (estudio D), que utiliza la información del estudio G para diseñar un procedimiento de medición que minimice el error para un propósito particular

y que define el universo al que se desea generalizar que puede contener algunas o todas las facetas y sus niveles en el universo de observaciones admisibles (Webb y Shavelson, 2005). Este estudio permite estimar los índices de generalizabilidad relativos y absolutos asociados, que asumen valores entre 0 y 1 funcionando de manera similar a un coeficiente de confiabilidad (Espinosa-Vázquez et al., 2017).

En la presente investigación se describe el diseño de un estudio de calidad percibida en servicios deportivos utilizando la herramienta CECASDEP (Gálvez-Ruiz, 2011), así como el análisis de la fiabilidad de los resultados a través de la teoría de la generalizabilidad, con el propósito de identificar las fuentes de variación que influyen en la evaluación de la calidad de manera que se pueda mejorar la aplicabilidad en futuras evaluaciones, así como también de obtener las estructuras numéricas generalizables a otras situaciones con características muestrales similares.

Método

Participantes

La recogida de información se realizó en dos instalaciones deportivas municipales de la provincia de Málaga con un total de 537 participantes (edad 32.11 ± 11.11 años). De ellos, 232 fueron mujeres (43.2%) con una edad media de 34.80 años (DT: 10.85) y un rango de edad desde 14 a 69 años. En el caso de los hombres, participaron 295 (54.9%) con un rango de edad desde 14 a 63 años y una edad media de 29.91 años (DT: 10.85). Se registraron 10 casos que no respondieron al género (1,9%) y en el caso de la edad fueron 21 valores perdidos, de los que 17 fueron hombres (5,8%) y 4 mujeres (1,7%). En relación a la formación académica, el 18.9% ($n=101$) tenían formación elemental o de primaria, el 35.7% tenían secundaria ($n=192$), y el 36.8% ($n=197$) contaban con formación universitaria, máster o doctorado, encontrándose un total de 47 participantes (8.8%) que no respondieron a esta pregunta. Finalmente en relación a la permanencia en el servicio deportivo, el 27.4% ($n=147$) llevaba como usuario menos de un año, el 15.8% ($n=185$) entre uno y dos años, el 12.3% ($n=66$) entre dos y tres años, obteniéndose porcentajes similares de usuarios con una permanencia de 44 (8.4%; $n=45$) y 5 años (7.8%; $n=42$). A partir de esa antigüedad los porcentajes bajan al 3%.

Material

El instrumento utilizado para este estudio recogió datos demográficos (género, edad, lugar de nacimiento, nivel de estudios y situación laboral), características de los clientes (actividad practicada, horario y permanencia en el centro), además de una medida sobre la percepción de calidad de

servicio (CECASDEP; Gálvez-Ruiz, 2011) compuesta por cinco dimensiones: *instalación municipal deportiva* (10 ítems), *espacio de actividad* (10 ítems), *vestuario* (12 ítems), *programa de actividades* (9 ítems) y *técnico deportivo* (10 ítems). Todos los ítems se evaluaron a través de una escala Likert de cinco puntos (1 = *nada de acuerdo*, 5 = *muy de acuerdo*). Los valores de fiabilidad (α) obtenidos para cada dimensión fueron satisfactorios y el análisis factorial confirmatorio mostró un ajuste adecuado de las distintas dimensiones ($c^2/df < 3$; CFI , GFI y $TLI > 0,90$; $RMSEA < 0,08$; $RMR < 0,05$) (Arbuckle, 2008; Kline, 2005; Worthington y Whittaker, 2006).

Procedimiento

Se empleó un muestreo no probabilístico por conveniencia, invitando a participar a 4 centros deportivos municipales con los que se había contactado previamente con el objetivo de mostrar la herramienta y explicar los objetivos de la investigación, aceptando formar parte del estudio 2 de ellos. Para la recogida de información colaboró un equipo de encuestadores formado por un total de 14 personas, todos alumnos en la Facultad de Psicología de la Universidad de Málaga cuya participación fue totalmente voluntaria. Se realizaron cuatro reuniones para lograr una adecuada familiarización con la herramienta de recogida de información. En este sentido, se revisaron los ítems del cuestionario con el objetivo de aclarar cualquier duda y se dieron las pautas necesarias para una correcta atención a los usuarios en caso de necesitar ayuda en la cumplimentación del cuestionario. El proceso de recogida de información tuvo duración de dos semanas completas y se utilizaron un total de 9 programas de actividad física ofertados por la organización con usuarios tanto de turno de mañana como de tarde, excluyendo para este estudio los programas de educación física especial y natación para niños (edad <10 años). Los participantes respondieron voluntaria y anónimamente al cuestionario en los diferentes espacios en los que se desarrollaban las actividades, disponiendo de 15 minutos para ello. Los instructores de las diferentes actividades prestaron su colaboración indicando a los usuarios la importancia de participar en el estudio para contribuir a la mejora del servicio prestado por la organización, finalizando las actividades dirigidas varios minutos antes para facilitar que los usuarios cumplimentaran el cuestionario.

Análisis estadístico

El análisis a través de la teoría de la generalizabilidad se compuso de 2 partes. En primer lugar, para el análisis de variabilidad en el estudio G se consideró un diseño multifaceta: usuarios (U), actividades (A), escalas (E) e ítems (I), siendo el diseño utilizado $U \times A \times E \times I$ (I:E), debido a que la faceta ítems está anidada en las escalas. En segundo lugar, a partir de los resultados del estudio G, se realizó el estudio D para obtener los índices relativos y absolutos, comprobando la variación de los índices al modificar los niveles originales de cada faceta.

Se utilizaron dos planes de medida diferentes, I:EE/AU donde *actividades* y *usuarios* constituían las facetas de generalización, pasando a ser las facetas de diferenciación en el plan de medida AU/I:EE, determinando por un lado cuál sería el número mínimo necesario de usuarios por actividad, y por otro comprobando que las escalas compuestas por un número diferente de niveles cada una (ítems) miden constructos diferentes, siendo por tanto mutuamente excluyentes.

El análisis de datos se realizó mediante el paquete estadístico SPSS v.21.0 (Statistical Package for the Social Sciences) y el software AMOS v.21.0 para el análisis factorial confirmatorio. El análisis de variabilidad se obtuvo mediante el programa SAS System v.9.1 (Scholtzhauer y Littell, 1997; SAS Institute Inc., 1999) y el desarrollo del análisis de generalizabilidad se realizó a través de EduG v.6.0 (Cardinet et al., 2010).

Resultados

En el plan de medida I:EE/AU, dividimos aleatoriamente la muestra con el objetivo de comprobar la fiabilidad y consistencia de la estructura numérica (N1=268; N2=269). La estimación de los componentes de varianza mostró que el diseño multifaceta *Usuarios x Actividades x Escalas x Ítems* fue significativo (<0.001), revelando que un bajo porcentaje de la variabilidad se asoció a las facetas *Escalas* (5.6% y 7.5%), *Usuarios* (13.6% y 13.3%), e *Ítems* (14.1% y 14.8%), mientras que la mayor parte de la variabilidad se asoció a la interacción *Usuarios/Ítems* (52.2% y 48,7%) (tabla 1). A continuación se realizó el estudio D (*decisión*) (Cronbach et al., 1972) o plan de optimización (Cardinet y Tourneur, 1978, 1985), recogiendo los valores de precisión para los coeficientes relativos y absolutos de generalizabilidad, realizando una optimización del plan de medida con los niveles originales mediante la modificación de las facetas de generalización.

Tabla 1. Resultado del análisis de variabilidad (diseño multifaceta I:EE/AU).

Fuente	dl		Suma de cuadrados		Cuadrado de la media		%	
	N1	N2	N1	N2	N1	N2	N1	N2
U	267	268	4179	4563	15.651	17.026	13.6	13.3
A	8	9	1	1	0.125	0.111	0	0
I:E	250	250	3529	4044	14.116	16.176	14.1	14.8
E	4	4	1198	1733	299.500	433.250	5.6	7.5
UE	1068	1072	3108	3632	2.919	3.388	14.5	15.6
UI:E	66750	67000	12830	13111	0.192	0.195	52.2	48.7

En la fase de optimización se modificaron los niveles originales del plan de medida hasta lograr una proyección sobre la estimación más adecuada de cada una de las facetas. En el plan de medida I:EE/AU, dividimos la muestra con el objetivo de comprobar la fiabilidad y consistencia de la estructura numérica. En este sentido, la tabla 2 muestra

los resultados de la primera mitad (268 usuarios) con un coeficiente de generalizabilidad excelente ($\xi\rho^2(\delta)$: 0.987; $\xi\rho^2(\Delta)$: 0.984), que observamos aumenta ligeramente a medida que añadimos más usuarios. La optimización de esta faceta se realiza aumentando el número de usuarios pues hemos considerado solamente la mitad de participantes.

Tabla 2. Optimización de la faceta usuarios en la primera mitad de la muestra en el plan de medida I:EE/AU.

Facetas	PLAN DE OPTIMIZACIÓN	I:EE/AU				
	Niveles originales	1	2	3	4	5
U	$n_u = 537$	268	280	300	325	350
A	$n_a = 9$	10	10	10	10	10
I	$n_i = 51$	51	51	51	51	51
E	$n_e = 5$	5	5	5	5	5
	Observaciones	615060	642600	688500	745875	803250
	$\xi\rho^2(\delta)$	0.987	0.988	0.988	0.989	0.990
	$\xi\rho^2(\Delta)$	0.984	0.985	0.986	0.987	0.988

La segunda mitad de la muestra (269 usuarios) obtiene un coeficiente de generalizabilidad ligeramente mejor ($\xi\rho^2(\delta)$: 0.989; $\xi\rho^2(\Delta)$: 0.987), si bien hay que tener en cuenta el mayor número de registros (685950). No obstante, en las dos mitades se comprueba que con una muestra inferior a

la original, concretamente del 50%, es posible obtener unos resultados muy satisfactorios (tabla 3), sin necesidad de aumentar la precisión de la generalización de la faceta pues aumentaría en un número muy grande el número de registros necesarios.

Tabla 3. Optimización de la faceta usuarios en la segunda mitad de la muestra en el plan de medida I:EE/AU.

Facetas	PLAN DE OPTIMIZACIÓN	I:EE/AU				
	Niveles originales	1	2	3	4	5
U	$n_u = 537$	269	280	300	325	350
A	$n_a = 9$	10	10	10	10	10
I	$n_i = 51$	51	51	51	51	51
E	$n_e = 5$	5	5	5	5	5
	Observaciones	685950	714000	765000	828750	892500
	$\xi\rho^2(\delta)$	0.989	0.989	0.990	0.991	0.991
	$\xi\rho^2(\Delta)$	0.987	0.987	0.988	0.989	0.990

La faceta *Actividades* también ofrece un coeficiente de generalizabilidad alto ($\xi\rho^2(\delta)$: 0.987; $\xi\rho^2(\Delta)$: 0.984). En este caso, se ha disminuido el número de actividades para comprobar si los coeficientes siguen siendo altos, a la vez que reducimos el número de registros de forma considerable.

De esta manera, comprobamos que los coeficientes se mantienen estables al reducir las actividades por lo que no resulta necesario utilizar todos los programas de actividad física ofertados, al obtener unos buenos resultados con 5 actividades (tabla 4) y pasando de 615060 a 341700 registros.

Tabla 4. Optimización de la faceta actividades en el plan de medida I:EE/AU.

Facetas	PLAN DE OPTIMIZACIÓN	I:EE/AU				
	Niveles originales	1	2	3	4	5
U	$n_u = 537$	268	268	268	268	268
A	$n_a = 9$	9	8	7	6	5
I	$n_i = 51$	51	51	51	51	51
E	$n_e = 5$	5	5	5	5	5
	Observaciones	615060	546720	478380	410040	341700
	$\xi\rho^2(\delta)$	0.987	0.987	0.987	0.987	0.987
	$\xi\rho^2(\Delta)$	0.984	0.984	0.984	0.984	0.984

En la tabla 5 se recogen los coeficientes de precisión de generalización en el plan de medida en el que la faceta *Escala* se situó, dentro del diseño, en el lugar de la generalización (o instrumentación), comprobando que los coeficientes de generalizabilidad relativos son altos pero los absolutos

muestran valores bajos ($\xi\rho^2(\Delta) < 0.90$), lo que permite constatar que aunque los registros son válidos, cada escala mide un constructo diferente al ser éstas mutuamente excluyentes.

Tabla 5. Optimización de la faceta escalas.

Facetas	PLAN DE OPTIMIZACIÓN	AU/I:EE				
	Niveles originales	1	2	3	4	5
U	$n_u = 537$	268	268	268	268	268
A	$n_a = 9$	9	8	7	6	5
I	$n_i = 51$	51	51	51	51	51
E	$n_e = 5$	5	6	7	8	9
	Observaciones	615060	738072	861084	984096	1107108
	$\xi\rho^2(\delta)$	0.814	0.840	0.859	0.875	0.887
	$\xi\rho^2(\Delta)$	0.760	0.791	0.816	0.835	0.850

Discusión y conclusiones

La TG puede utilizarse para el diseño de estudios con una optimización más eficiente, pues según Blanco-Villaseñor y Losada (2004) su aplicación resulta fructífera. Concretamente en el ámbito de la gestión deportiva, podemos conocer el tamaño adecuado de cada una de las variables que forman parte del diseño de medida para poder generalizar con precisión, y la importancia de la presente investigación recae en la escasez de estudios relacionados con la gestión y los servicios deportivos por lo que aporta una importante aplicabilidad y facilita una adecuada toma de decisiones en un sector en permanente cambio (Gálvez-Ruiz, Rosado y Romero-Galisteo, 2015).

La confiabilidad, como uno de los principales índices

que aportan evidencia de validez en los resultados de una evaluación (Downing y Yudkowsky, 2009), se ha analizado a través de la TG obteniéndose resultados que se asemejan a los obtenidos por Gálvez-Ruiz, Sánchez-Oliver, Baena-Arroyo y Blanco-Villaseñor (2016), donde la varianza atribuida a la interacción entre los usuarios y los ítems tras el estudio G fue la más elevada, concretamente del 49.9% que supone un resultado muy similar a los obtenidos en esta investigación (52.2% para N1 y 48.7% para N2). Si bien en el presente trabajo no se consideró el género como una faceta dentro del modelo tras comprobar la nula variabilidad en el estudio de Gálvez-Ruiz et al. (2016), sí se tuvieron en cuenta las actividades debido a los amplios programas de actividades ofertados por las organizaciones deportivas como parte de la propuesta de valor de las mismas. Sin embargo,

los resultados del estudio G han mostrado que al igual que ocurría con el género, ésta faceta no aporta variabilidad (0%) y consideramos que se trata de un hallazgo muy interesante para futuros estudios, puesto que en la configuración de un horario de actividades influye el perfil y volumen de usuarios en diferentes franjas horarias.

En cuanto al análisis de generalizabilidad, para el estudio D se realizó una estimación tanto del tamaño muestral adecuado como del número mínimo de programas de actividad física, para la cual se estimó una optimización del diseño de medida EI:EUA, constituyendo los *usuarios* y *actividades* las facetas de generalización. En este sentido, los resultados obtenidos en el estudio de generalizabilidad, nos permiten considerar adecuado el diseño de medida de partida, tanto en el número de participantes como de actividades consideradas. Sin embargo, existe la posibilidad de disminuir considerablemente el número de registros con la misma precisión de generalización en posteriores análisis. Concretamente, podemos reducir el número de usuarios sin que afecte a los coeficientes relativos y absolutos, incidiendo con ello en un menor número de registros. Así, podríamos emplear una muestra entre 250 y 300 participantes obteniendo altos valores de fiabilidad y generalizabilidad ($\xi\rho^2(\Delta)=0.98$), lo que resulta de gran utilidad teniendo en cuenta el elevado número de usuarios con los que cuentan las organizaciones de servicios deportivos actualmente. Los valores obtenidos en este estudio han sido muy similares a los obtenidos por Gálvez-Ruiz et al. (2016). En la evaluación de los programas de actividad física ofertados por la organización, hemos comprobado que se podrían reducir los niveles de la faceta obteniendo resultados igualmente fiables y generalizables. Sin embargo, al no aportar variabilidad al diseño de medida, futuros estudios deberán considerar la necesaria utilización de esta faceta.

Como conclusión, la importante aplicabilidad de la TG aporta a los resultados gran relevancia de cara a futuros estudios, en los que se pueda aportar mayor precisión de generalización utilizando el mínimo de recursos posibles. Este hecho adquiere gran relevancia dentro del ámbito de la gestión deportiva, al contar las organizaciones con altos volúmenes de usuarios y donde los programas de actividad física ofertados no sólo son diversos, sino que muchos de ellos cuentan con varios grupos de usuarios, haciendo la tarea de evaluación más lenta y, sobre todo, costosa. Por este motivo, el presente análisis de generalizabilidad debe constituir un estudio *a priori* que nos sirva como base para el diseño de una investigación de mayor alcance, en el que no sea necesaria la evaluación de todos los grupos de un mismo programa de actividad física, así como tampoco de todos los usuarios con los que cuenta la organización.

Limitaciones y futuras líneas

El desarrollo de nuevos modelos dentro del mercado de los servicios deportivos, como por ejemplo el *low-cost*, los centros *boutique* o el desarrollo de centros concesionales, han ampliado en un 3.2% el número de clubs a nivel europeo favoreciendo con ello el aumento de usuarios de servicios deportivos un 0.7% en el último año según datos proporcionados por el último European Health & Fitness Market Report (EuropeActive & Deloitte, 2018), ocupando España el quinto lugar en cuanto a volumen de usuarios a nivel europeo. Estos datos abren sin duda la posibilidad a continuar realizando estudios en esta línea de trabajo que faciliten adecuados instrumentos de evaluación y el conocimiento de todas las variables que determinan la calidad del servicio. Además, la importancia de la aplicación de la teoría de la generalizabilidad permitirá continuar con la optimización de los diseños de medida pues la rápida evolución que está experimentando el mercado de servicios deportivos así lo requiere.

En este sentido, la diversidad de modelos de negocio con diferentes propuestas de valor (distintas instalaciones, equipamientos, horarios, actividades dirigidas, atención al usuario, servicios especializados, etc.) requiere estudios adaptados a las características de los mismos, donde la teoría de generalizabilidad nos permitirá comprobar la variabilidad aportada por cada una de las facetas al modelo de medida en función de la tipología de modelo de servicio deportivo.

Por ello, ante la especificidad del instrumento la teoría de la generalizabilidad permite medir la confiabilidad de una prueba por medio de la cuantificación de la importancia de cada una de sus fuentes de variabilidad (Zúñiga-Brenes y Montero-Rojas, 2007) a través de las diferentes facetas y las relaciones entre ellas. Por otro lado, la evolución demográfica y poblacional también suponen aspectos a tener muy en cuenta en futuros estudios. Las organizaciones deben ser tan dinámicas como la economía y, por tanto, requieren conocer las características tanto de los clientes actuales como los potenciales, además de los cohortes generacionales con el objetivo de ofrecer el mejor servicio. Por ello, pensamos que futuros estudios deben considerar la edad como una variable de importancia, utilizando las generaciones establecidas por la International Health, Racquet & Sports club Association – IHRSa (2017) cuyos rangos de edad tienen presencia mayoritaria y coexisten en las instalaciones deportivas actualmente: Baby Boomers, Generación X, Generación Y o Millenials, y Generación Z. En esta línea, un reciente estudio ha utilizado estas generaciones mostrando las diferentes relaciones en las dimensiones que influyen en la cadena de fidelidad (García-Fernández, Gálvez-Ruiz, Sánchez-Oliver y Grimaldi-Puyana, en prensa), por lo que puede suponer una faceta de variabilidad más dentro de un modelo de medida teniendo en cuenta que en determinados modelos de negocio la muestra presenta

ciertas limitaciones y no es tan amplia como puede serlo en los modelos concesionales y *low-cost*, por ejemplo.

Sin embargo, este estudio tiene ciertas limitaciones siendo el tipo de muestreo (por conveniencia) una de las mismas pese a tratarse del método de reclutamiento más comúnmente empleado en este tipo de estudios. Sin embargo, la actual competencia existente en el sector con un mercado muy atomizado dificulta la posibilidad de obtención de información y acceso a muestras de usuarios. Por otro lado, los resultados obtenidos permiten generalizar los resultados dentro de un modelo de servicios deportivos municipales, y pese a que los

resultados del análisis de variabilidad mostró similitud con un estudio aplicado en centros de fitness (Gálvez-Ruiz et al., 2016), no podemos afirmar que los resultados obtenidos en el presente estudio sean generalizables en muestras procedentes de otras tipologías de servicios deportivos, aunque si debemos considerarla como información relevante a tener en cuenta a la hora de plantear futuros estudios. De hecho, futuras investigaciones deben partir de muestras con un tamaño de 300 participantes como mostró el estudio de Gálvez-Ruiz et al. (2016) y el presente trabajo, buscando la optimización de esta faceta a partir de este nivel.

Referencias

1. Arbuckle, J. L. (2008). *AMOS 17 user's guide*. Chicago, IL: SPSS.
2. Baker, D. A. y Crompton, J. L. (2000). Quality, satisfaction and behavioral intentions. *Annals of Tourism Research*, 27(3), 785-804.
3. Barros, C. y Gonçalves, L. (2009). Investigating individual satisfaction in health and fitness training centres. *International Journal of Sport Management and Marketing*, 5(4), 384-395.
4. Batista, J. M. y Coenders, G. (2012). *Modelos de ecuaciones estructurales*. Madrid: La Muralla.
5. Bitner, M. J., Faranda, W. T., Hubbert, A. R., y Zeithaml, V. A. (1997). Customer contributions and roles in service delivery. *International Journal of Service Industry Management*, 8(3), 193-205.
6. Blanco-Villaseñor, Á. (1991). La teoría de la generalizabilidad aplicada a diseños observacionales. *Revista Mexicana de Análisis de la Conducta*, 17(3), 23-63.
7. Blanco-Villaseñor, Á. (2001). Generalizabilidad de observaciones uni y multifaceta: estimadores LS y ML. *Metodología de las Ciencias del Comportamiento*, 3(2), 161-193.
8. Blanco-Villaseñor, Á. y Losada, J. L. (2004). Generalización en Diseños Observacionales: alternativas de estimación y modelización. *Metodología de las Ciencias del Comportamiento (suplemento)*, 83-88.
9. Blanco-Villaseñor, Á. y Morales-Sánchez, V. (2010). Teoría de la generalizabilidad: investigaciones aplicadas. En M. J. Blanca, R. Alarcón y D. López-Montiel (Coords.), *XI Congreso de Metodología de las Ciencias Sociales y de la Salud. Libro de Actas* (142). Málaga: UMA-Tecnolex.
10. Cardinet, J., Johnson, S. y Pini, G. (2010). *Applying Generalizability Theory using EduG*. London: Routledge.
11. Cardinet, J. y Tourneur, Y. (1978). *Le calcul des marges d'erreur dans la théorie de la généralisabilité*. Mons: Université de l'Etat, Service d'étude des méthodes et des moyens d'enseignement.
12. Cardinet, J. y Tourneur, Y. (1985). *Assurer la mesure*. Berne: Peter Lang.
13. Cronbach, L. J., Gleser, G. C., Nanda, H. y Rajaratnam, N. (1972). *The dependability of behavioral measurements: theory of generalizability for scores and profiles*. Nueva York: John Wiley and Sons.
14. Cronin, J. J. y Taylor, S. A. (1992). Measuring service quality: A reexamination and extension. *Journal of Marketing*, 56, 55-68.
15. Downing, S. M. y Yudkowsky, R. (2009). Reliability. En S. M. Downing y R. Yudkowsky (Eds.), *Assessment in Health Professions Education* (pp. 57-74). Nueva York: Routledge.
16. Espinosa-Vázquez, O., Martínez-González, A., Sánchez-Mendiola, M. y Leenen, I. (2017). Análisis de un examen clínico objetivo estructurado en odontología desde la teoría de la generalizabilidad. *Investigación en Educación Médica*, 6(22), 109-118.
17. EuropeActive & Deloitte (2017). *European Health & Fitness Market Report*. Cologne: EuropeActive.
18. EuropeActive & Deloitte (2018). *European Health & Fitness Market Report*. Cologne: EuropeActive.
19. Gálvez-Ruiz, P. (2011). *Cuestionario para evaluar la calidad de servicios deportivos: estudio inicial de las propiedades psicométricas*. Málaga: Servicio de Publicaciones de la Universidad de Málaga.
20. Gálvez-Ruiz, P., Rosado, A. y Romero-Galisteo, R. P. (2015). Validación de la versión reducida del CECASDEP en usuarios de servicios deportivos. *Suma Psicológica*, 22(2), 78-85.
21. Gálvez-Ruiz, P., Sánchez-Oliver, A. J., Baena-Arroyo, M. J. y Blanco-Villaseñor, Á. (2016). Estimación muestra en la percepción de los servicios deportivos a través de la teoría de la generalizabilidad. *Revista de Psicología del Deporte*, 25(Suppl. 1), 65-68.
22. García-Fernández, J., Gálvez-Ruiz, P., Sánchez-Oliver, A. J. y Grimaldi-Puyana, M. (en prensa). Customer loyalty in fitness centres: differences according to Millennials, Baby Boomers, and Generation X and Z. En B. G. Pitts y J. J. Zhang (Eds.), *Global Sport Management. Contemporary issues and inquiries*. New York, NY: Routledge.
23. García-Fernández, J., Gálvez-Ruiz, P., Vélez-Colón, L., Ortega-Fernández, J. y Fernández-Gavira, J. (2018). Exploring fitness centre consumer loyalty: differences of non-profit and low-cost business models in Spain. *Economic Research*, 31(1), 1042-1058.
24. Gocłowska, S. y Piatkowska, M. (2017). Service Satisfaction and sport consumption in the fitness center in Warsaw. *European Journal of Service Management*, 22(2), 31-37.
25. International Health, Racquet & Sports club Association (2017). *The IHRSA Global Report 2016*. Boston: IHRSA.
26. ISO 10075-3 (2004). *Ergonomic principles related to mental work-load. Part 3: Principles and requirements concerning methods for measuring and assessing mental workload*. Geneva: ISO.
27. Jiménez, R., Moreno, B., Leyton, M. y Claver, F. (2015). Motivación y estadios de cambio para el ejercicio físico en adolescentes. *Revista Latinoamericana de Psicología*, 47(3), 196-204.
28. Kline, R. (2005). *Principles and practice of structural equation modeling* (2ª ed.). Nueva York, NY: The Guilford Press.
29. Parasuraman, A., Zeithaml, V., y Berry, L. (1985). A conceptual model of service quality and its implications for future research. *Journal of Marketing*, 49, 41-50.
30. SAS Institute Inc. (1999). *SAS/STAT User's Guide*, v7.1. Cary, NC: SAS Institute Inc.
31. Schlotzhauer, S. D. y Littell, R. C. (1997). *SAS System for Elementary Statistical Analysis* (2ª Ed.). Cary, NC: SAS Institute Inc.
32. Theodorakis, N. D., Howat, G., Ko, Y. J. y Avourdiadou, S. (2014). A comparison of service evaluation models in the context of sport and fitness centres in Greece. *Managing Leisure*, 19(1), 18-35.
33. Webb, N. M. y Shavelson, R. J. (2005). Generalizability Theory: Overview. En B. S. Everitt y D. C. Howell (Eds.), *Encyclopedia of Statistics in Behavioral Science* (Vol. 2) (pp. 717-719). Chichester: John Wiley &

Sons.

34. Wickel, E. E. y Welk, G. J. (2010). Applying Generalizability Theory to estimate habitual activity levels. *Medicine and Science in Sports and Exercise*, 42, 1528-1534.
35. Worthington, R. y Whittaker, T. (2006). Scale development research. A content analysis and recommendations for best practices. *The Counselling Psychologist*, 34(6), 806-838.
36. Yi, Y., Gong, T., y Lee, H. (2013). The impact of other customers on customer citizenship behavior. *Psychology and Marketing*, 30(4), 341-356.
37. Zúñiga-Brenes, M. E. y Montero-Rojas, E. (2007). Teoría G: un futuro paradigma para el análisis de pruebas psicométricas. *Actualidades en Psicología*, 21, 117-144.