REVISTA INTERNACIONAL DE DERECHO Y GESTIÓN DEL DEPORTE

International Journal of Sports Law & Management

ANÁLISIS DE DATOS EN GESTIÓN DEPORTIVA: LA CONMENSURABILIDAD ENTRE TAMAÑOS DE EFECTO PROVENIENTE DE ESCALAS DE MEDIDA CON DIFERENTE RANGO

José Antonio Martínez García
Universidad Politécnica de Cartagena

Laura Martínez Caro
Universidad Politécnica de Cartagena

RESUMEN

Este trabajo plantea una cuestión no tratada hasta ahora en el análisis de datos en investigación en gestión deportiva: si los tamaños de efecto provenientes de diferentes estudios que utilicen escalas tipo Likert de distinto rango son conmensurables. El índice susceptible de análisis es la diferencia de medias estandarizada. Para ello, se deducen las expresiones que relacionan la posible distorsión propiciada por el cambio de escala sobre las medias y la varianza observada intra-grupos. El análisis de esas ecuaciones permite llegar a las siguientes conclusiones: (1) Escalas de diferente rango producen inevitablemente una distorsión de la medida del fenómeno; (2) Los tamaños de efecto son conmensurables siempre que haya invarianza de escala, lo que sucede cuando esa distorsión es mínima; (3) Cuando no se puede asegurar la invarianza de escala, la comparación entre estudios puede enmascarar diferencias cuando realmente las haya o, por el contrario, reportar diferencias cuando efectivamente no existan.

ABSTRACT

This study analyses the problem of commensurability between effect sizes calculated from Likert scales with different rank. Standardized mean difference has been the index considered for that purpose. We have defined the scale invariance term and we have deduced the expressions that relate the possible distortion provoked by changing the scale rank on the mean and variance statistics. The analysis of these equations drives to the following conclusions: (1) Different rating scales distort measurement; (2) Effect sizes are commensurable to the extent that scale invariance holds. This is achieved when distortion is negligible; (3) When scale invariance is not guaranteed, the comparison between different studies can yield misleading results.

PALABRAS CLAVE: Conmensurabilidad, escalas de medida, tamaños de efecto, gestión deportiva, análisis de datos.

KEY WORDS: Conmensurability, ratings cales, effect sizes, sport management, data analysis.

Fecha de recepción: 28/12/2008 Fecha de aceptación: 22/01/2009

INTRODUCCIÓN

La investigación en gestión deportiva se está acercando paulatinamente a la rigurosidad analítica característica de otras disciplinas, como la psicología o la economía. Así, por ejemplo, existen estudios que ofrecen pautas de actuación sobre cómo realizar ciertos análisis de datos, en aspectos básicos como la detección e interpretación de efectos[1] y el tratamiento de la invarianza de escala[2].

Dado que aún no existe consenso en las ciencias sociales sobre la arbitrariedad o no de las medidas de los fenómenos de interés, el rango óptimo que deben tener las escalas propuestas para las mediciones, o la consideración sobre el tratamiento continuo o discreto de las medidas provenientes de escalas ordinales, existe una multiplicidad de enfoques que se traducen en una heterogeneidad metodológica. Aunque, a priori, estas divergencias pueden enriquecer globalmente la disciplina, no es menos cierto que suponen una gran dificultad a la hora de homogeneizar y resumir los resultados de las investigaciones.

Desde hace más de 30 años, el metanálisis se ocupa de sistematizar los resultados empíricos con el fin de contribuir a la acumulación de conocimiento. A las dificultades provenientes de esa heterogeneidad de perspectivas, se suman, entre otros factores, una serie de artefactos estadísticos que afectan a la magnitud de los efectos encontrados[3]. Sin embargo, entre todas esas amenazas a la validez del metanálisis, no hemos encontrado ninguna referencia al problema de homogeneizar efectos provenientes de diferentes escalas de medida. Considerando el uso generalizado de las escalas tipo

[1] PARKS, J. B.; SHEWOKIS, P. A.; COSTA, C. A. "Using statistical power analysis in sport management research". *Journal of Sport Management*. 1999, vol. 13, núm. 2, p. 139-147.
[2] MARTÍNEZ, J. A. "Estudio de la invarianza de escala mediante el método de cálculo integral en la medición de la calidad percibida de los servicios deportivos". *Revista Internacional de Ciencias del Deporte*. En prensa.
[3] HUNTER, J. E.; SCHMIDT, F. L. *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*. 2nd Edition. Newbury Park: Sage Publications, 2004.

Likert en la investigación deportiva (y en todas las áreas de la investigación social), estimamos preceptivo el profundizar sobre la naturaleza de esta cuestión.

El objetivo de este trabajo es, de este modo, examinar si los tamaños de efecto calculados a partir de estudios que utilizan diferentes rangos de escalas son conmensurables, es decir, si la posible distorsión producida por la diferencia de rango en el instrumento de medida afecta a la magnitud del efecto encontrado, y si ello permite esa comparación. El desarrollo de nuestro razonamiento será especificado analíticamente en relación a la definición de invarianza de escala. Asimismo, se discutirá la relación del Factor de Imprecisión sobre la Escala de Medida[1], *FIEM*, con la varianza observada de los estudios. La diferencia de medias tipificada, correspondiente a la "familia *d*" de tamaños de efecto, será el índice susceptible de análisis.

LA INVARIANZA DE ESCALA

La invarianza de escala se define 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 como $g(\sum_{i=1}^{n} f(x)) = S^2$, donde

[1] MARTÍNEZ, J. A.; MARTÍNEZ, L. "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*. 2008, vol. 20, núm. 2, p. 305-310.

$f(x) = ax^k$, $x = (x_i - \overline{x})$, $k=2$ 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[1] en el intervalo [0,1]. De este modo, se puede establecer una comparación directa entre todas las respuestas, existiendo diversos procedimientos para ello[2], como la comparación entre medias, los modelos de ecuaciones estructurales, el análisis de entropía o el cálculo de la diferencia de área entre dos funciones.

Parece evidente que si existe invarianza de escala, el tamaño de efecto *d*, o diferencia de medias estandarizada, no debería de verse afectado. La diferencia estandarizada entre grupos es (1):

$$d_1 = \frac{M_{a1} - M_{b1}}{S_1} \qquad (1)$$

siendo M_{a1} y M_{b1} las medias de los dos grupos y S_1 la desviación típica intra-grupos que se puede obtener mediante una ponderación de las desviaciones de los dos grupos, si existe homogeneidad de varianza (2):

$$S_1 = \sqrt{\frac{(n_{a1}-1)S_{a1}^2 + (n_{b1}-1)S_{b1}^2}{n_{a1}+n_{b1}-2}} \qquad (2)$$

siendo n_{a1} y n_{b1} los tamaños muestrales de los grupos y S_{a1}^2 y S_{b1}^2 las varianzas insesgadas.

[1] COHEN, P. et al. “The problem of units and the circumstance for POMP”. *Multivariate Behavioral Research*. 1999. vol. 34, p. 315-346.

[2] MARTÍNEZ, J. A. “Estudio de la invarianza de escala mediante el método de cálculo integral en la medición de la calidad percibida de los servicios deportivos”. *Revista Internacional de Ciencias del Deporte*. En prensa.

Si existe invarianza de escala, entonces $M_{a1} = cM_{a2}$, $M_{b1} = cM_{b2}$, y $S_1 = cS_2$, con lo que el tamaño de efecto permanece constante.

Del mismo modo, dos escalas distintas invariantes deberían ser igual de imprecisas. El Factor de Imprecisión sobre la Escala de Medida (*FIEM*) es invariante frente a los cambios de escala. Hay que recordar que los errores absoluto y relativo están en función de la escala de medición, pero no así *FIEM*, que es independiente. *FIEM* se obtiene de la siguiente expresión (3).

$$FIEM = \left(\frac{z}{R_E} \frac{S}{\sqrt{n}} \right) \tag{3}$$

donde *z* es el valor crítico de la distribución normal estándar para un nivel de confianza determinado (normalmente 0.05), y *S* es la desviación típica muestral.

Podemos, asimismo, construir un valor de *FIEM* medio ($FIEM_{M1}$) para los dos grupos dentro de la muestra utilizando la desviación típica intra-grupos, bajo el supuesto de homogeneidad (4):

$$FIEM_{M1} = \left(\frac{z}{R_{E1}} \frac{S_1}{\sqrt{n_1}} \right) \tag{4}$$

De nuevo es fácil comprobar cómo, para la medición del mismo fenómeno con otra escala distinta, $FIEM_{M2}= FIEM_{M1}$ siempre que se mantenga el mismo nivel de confianza y los tamaños muestrales sean iguales.

En resumen, si un fenómeno es invariante frente a la escala de medida, no existirá distorsión en el tamaño de efecto calculado más allá del correspondiente a la variabilidad muestral, las medias y las varianzas guardarán una relación de proporcionalidad, y las imprecisiones de las estimaciones de las medias serán también equivalentes.

EL EFECTO DEL CAMBIO DE ESCALA

El procedimiento de conversión de cualquier escala a una puntuación en el intervalo (0,1), obedece a la siguiente expresión (5):

$$x_i^{'} = \left[\left(\frac{x_i - L_{\inf}}{R_E} \right) \right] \qquad (5)$$

donde $x_i^{'}$ es el valor convertido al intervalo (0,1), y x_i es el valor de la puntuación de cada unidad muestral para ambas escalas y $L_{\inf}$ es el extremo inferior del rango de escala - R_E - considerado.

Es importante resaltar que cuando hemos hablado de la conversión de valores por el factor $c = R_{E2} / R_{E1}$, los valores convertidos no tienen por qué concordar exactamente con los de la nueva escala, ya que esa nueva escala puede tener o no el mismo origen. Es decir, por ejemplo, para la conversión de valores de una escala de 1 a 5 en una escala de 1 a 7, la multiplicación por *c* produciría valores en la segunda escala fuera de los límites del extremo superior del rango. Es por ello, que podemos afirmar que realmente los cambios de escala obedecerían a otra expresión en cuanto a sus valores (6):

$$x_{i2} = \left[x_{i1} \left(\frac{R_{E2}}{R_{E1}} \right) \right] - \left(\frac{L_{\sup 2} - L_{\sup 1}}{R_{E1}} \right) \qquad (6)$$

donde x_{i2} y x_{i1} son los valores de la puntuación de cada unidad muestral para ambas escalas y $L_{\sup 2}$ y $L_{\sup 1}$ son los valores superiores de los rangos de escalas considerados. Es decir, si el estudio 1 se realiza en una escala de 1 a 5 y el estudio 2 en una escala de 1 a 7, los valores de L_2 y L_1 serían 7 y 5. Esta conversión produce que el valor de la puntuación de la escala 2 no sea un valor entero para ciertos casos.

La expresión (6) se modifica cuando las escalas con límite inferior igual a 1, se transforman escalas con origen en 0, por ejemplo, para pasar de una escala de 1 a 5 a una escala de 0 a 10 (7):

$$x_{i2} = \left[x_{i1} \left(\frac{R_{E2}}{R_{E1}} \right) \right] - \left(\frac{R_{E2}}{R_{E1}} \right) \tag{7}$$

Como puede contemplarse en la Tabla 1, la conversión produce distorsión en los valores transformados, ya que algunos de esos valores no corresponde con valores enteros de la nueva escala, por lo que el individuo tendría que elegir entre los dos valores enteros más cercanos a su correspondiente valor decimal. Evidentemente, ello podría suponer una deformación de los valores medios y de la varianza de la distribución de datos.

Tabla 1. Conversión de los valores de una escala a otra tras aplicar (6) y (7)

Conversión escala de 1 a 5 en escala de 1 a 7		Conversión escala de 1 a 7 en escala de 1 a 11		Conversión escala de 1 a 7 en escala de 0 a 10	
1	1	1	1	1	0
2	2.5	2	2.66	2	1.66
3	4	3	4.33	3	3.33
4	5.5	4	6.00	4	5.00
5	7	5	7.66	5	6.66
		6	9.33	6	8.33
		7	11	7	10

No obstante, es mucho más cómodo seguir operando con *c*, como factor de conversión ya que, aunque produzca valores fuera del intervalo considerado, la diferencia de medias entre grupos no se vería afectada, así como tampoco la varianza de la distribución. Sin embargo, los resultados comentados en la Tabla 1, nos indican cómo se produce una inevitable distorsión de la información cuando se cambia de escala. Por ejemplo, si suponemos que una escala de 1 a 7 es la que mejor refleja la escala mental que un individuo construye para valorar un fenómeno, y se le muestra un cuestionario con una escala de 0 a 10, la media y la varianza de las puntuaciones muy posiblemente no se correspondan perfectamente con una transformación invariante de la puntuación que habrían obtenido utilizando una escala de 1 a 7. Este hecho hace que se añada un factor contaminante proveniente del tipo de escala utilizado, que complica el tratamiento estadístico de la invarianza, y por ende, afecta al análisis de la conmensurabilidad de tamaños de efecto, como veremos a continuación.

LA CONMENSURABILIDAD DE LOS TAMAÑOS DE EFECTO

Suponemos el caso genérico en el que un estudio se replica en una población distinta y con una escala Likert de diferente rango. El valor de la media del grupo a M_{a2} se puede descomponer en la siguiente expresión (8):

$$M_{a2} = (M_{a1} + q_{a2})c + p_{a2} \tag{8}$$

siendo p_{a2} y q_{a2} la distorsión en el valor medio producido por el cambio de escala y por la nueva población, respectivamente, y que pueden tener valores positivos o negativos. De forma análoga (9):

$$M_{b2} = (M_{b1} + q_{b2})c + p_{b2} \tag{9}$$

Del mismo modo, la desviación intra-grupos del segundo estudio también puede descomponerse (10):

$$S_2 = (S_1 + Q_2)c + P_2 \tag{10}$$

siendo P_2 y Q_2 la distorsión en la desviación observada en el segundo estudio debido al efecto de la escala y a la recogida de datos en una nueva población, respectivamente. Pueden ser valores positivos o negativos.

Si la magnitud del efecto real en las dos poblaciones fuera el mismo, es decir $d_1=d_2$, y existiera invarianza de escala, bajo la asunción de inexistencia de ningún otro factor moderador, se debería cumplir la siguiente igualdad (11):

$$\frac{M_{a1} - M_{b1}}{S_1} = \frac{\left[(M_{a1} - M_{b1}) + (q_{a2} - q_{b2})\right]c + (p_{a2} - p_{b2})}{(S_1 + Q_2)c + P_2} \tag{11}$$

Rápidamente se deduce que P_2 , p_{a2}, p_{b2} deberían ser cero o muy próximos a cero. Siguiendo con el razonamiento que mostrábamos al comienzo del trabajo, la diferencia de medias y la desviación típica deberían ser estadísticamente iguales para ambos estudios.

Sin embargo, la expresión (11) nos muestra más información, si relajamos la asunción de invarianza de escala. Los tamaños de efecto d_1 y d_2 también serían iguales si (12):

$$\frac{(p_{a2} - p_{b2})}{P_2} = d_1 \qquad (12)$$

Estas deducciones nos conducen a postular que los tamaños de efecto provenientes de dos estudios en poblaciones distintas, y medidos con escalas de diferente rango, pueden ser reportados como iguales cuando realmente existan diferencias en esa segunda población, ya que el cambio de escala podría enmascarar esa divergencia. Es decir, podría haber cambios importantes en la desviación típica o en la diferencia de medias, o en ambos casos, y que ello no modificase el tamaño de efecto por una compensación de la distorsión de la escala.

Es evidente que si suponemos que los tamaños de efecto son distintos entre las dos poblaciones, podríamos tener el problema inverso, es decir, que empíricamente fueran iguales debido al efecto de la escala.

Si consideramos ahora, que nos movemos en diferentes estudios dentro de la misma población, podemos asumir entonces que Q_2 , q_{a2}, q_{b2} son cero. Se puede demostrar que cuando $P_2 = 0$, $FIEM_{M2}= FIEM_{M1}$ si $n_1=n_2$, es decir, los estudios tienen la misma precisión, ya que realizando unas simples operaciones algebraicas, se deriva la siguiente expresión (13):

$$P_2 = \frac{R_{E2}}{z}\left(FIEM_{M2}\sqrt{n_2} - FIEM_{M1}\sqrt{n_1}\right) \qquad (13)$$

Esto quiere decir que, en dos estudios en la misma población con tamaños muestrales iguales y que sean equivalentes en su imprecisión, podrá considerarse que existe invarianza de escala, siempre que se asuma que la escala no produce ningún efecto sobre las medias de los grupos ($p_{a2}=p_{b2}=0$) o que de existir distorsión de la escala sobre los valores medios, es equivalente entre grupos, es decir, $p_{a2}=p_{b2}$.

De este modo, a través del uso de *FIEM*, podríamos saber si P_2 está próximo a cero, lo que nos daría información sobre la importancia del efecto de la escala sobre la varianza observada. Asimismo, otra opción adecuada sería el uso del contraste de varianzas a través del estadístico *F*-Snedecor. Para ello, se re-escalaría la varianza del primer estudio por el factor c^2 y se compararía con la varianza del segundo estudio. Si ambas varianzas no son estadísticamente diferentes, es muy posible que el efecto de P_2 sea prácticamente nulo.

En un reciente estudio[1], tras recoger datos sobre tres muestras aleatorias provenientes de la misma población, se muestra empíricamente cómo las escalas Likert de 1 a 5, 1 a 7, y 1 a 10 no producen cambios importantes sobre la media y la varianza. Estos resultados apoyan el hecho de que la medición de un fenómeno en el mismo contexto por escalas de diferente rango es estadísticamente invariante (aunque sería preferible realizar una triangulación estadística con otros métodos para confirmarlo). Es decir, el efecto de la distorsión de la escala es tan pequeño, que no es capaz de desviar la varianza observada más allá de la desviación esperada por la variabilidad muestral. Sin embargo, esta aseveración puede ser en cierto modo arriesgada, ya que podría haber diferencias importantes entre las varianzas que quedaran mitigadas por esa distorsión de la escala (P_2 puede ser positivo o negativo). Por tanto, no tenemos completa certeza acerca de la

[1] DAWES, J. “Do data characteristics change according to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales”. *International Journal of Market Research*, 2008. vol. 50, núm. 1, p. 61-77.

invarianza estadística ya que no se separa la distorsión de la variabilidad muestral de la distorsión del efecto de la escala.

DISCUSIÓN

En este trabajo hemos planteado la cuestión de la conmensurabilidad de tamaños de efecto entre estudios que utilizan escalas de medida tipo Likert con diferente rango. De la deducción realizada se derivan varias consecuencias:

En primer lugar, podemos afirmar que las escalas de medida con diferente rango producen inevitablemente una distorsión en la medida del fenómeno en cuestión, derivada de la propia especificación de la escala, ya que una transformación invariante perfecta es imposible por la no concordancia de los valores decimales.

A la distorsión producida por la especificación de la escala, habría que añadir el posible efecto de la sensibilidad del individuo hacia la escala, es decir, puede que los encuestados se sientan más cómodos utilizando escalas de 0 a 10 que de 1 a 5, por ejemplo.

En cualquier caso, la cuestión relevante es si esa distorsión es importante o no. Para analizar este hecho, proponemos analizar la invarianza de escala tal y como se ha especificado en otros estudios[1]. Reconocemos, asimismo, que este método podría producir también un sesgo, ya que algunos sujetos podrían tratar de realizar sus valoraciones en función de su respuesta en la primera escala, por lo que existiría dependencia en las respuestas sucesivas. Para tratar de mitigar esta limitación, sería

[1] MARTÍNEZ, J. A. "Estudio de la invarianza de escala mediante el método de cálculo integral en la medición de la calidad percibida de los servicios deportivos". *Revista Internacional de Ciencias del Deporte*. En prensa.

recomendable diseñar estudios donde se comparen sólo dos escalas, aleccionando bien a los individuos sobre cómo deben realizar la valoración, o separando lo más posible ambas valoraciones dentro del cuestionario (si es que el estudio tiene más objetivos).

En segundo término, hemos deducido cómo en estudios realizados sobre la misma población utilizando diferentes muestras, no se puede aislar el efecto de la escala de la variabilidad inherente al muestreo, aunque bien es cierto que la utilización de la expresión (13) o del test *F*-Snedecor, sería una buena aproximación al estudio de la invarianza escalar. Sin embargo, la comparación entre tamaños de efecto es más interesante entre diferentes poblaciones, es decir, en la replicación de estudios con diferentes situaciones de test o contextos diferentes[1]. Para una misma población o contexto determinado, se podría construir un intervalo de confianza para el tamaño de efecto proveniente de una única muestra[2] por lo que ya dispondríamos de información acerca de la variabilidad del índice, sin necesidad de analizar otras muestras de esa población.

Por último, hemos demostrado cómo para el caso de conmensurar tamaños de efecto de estudios realizados en diferentes poblaciones, que por definición pueden estar sujetos a cambios en esa magnitud del efecto, sólo es posible la conmensurabilidad si existe invarianza de escala. Cuando no podemos asegurar esa invarianza, las expresiones (11) y (12) nos indican que la comparación entre estudios puede enmascarar diferencias cuando realmente las haya o, por el contrario, reportar diferencias cuando efectivamente no existan.

Finalmente, los investigadores en gestión deportiva deberán tener en cuenta estas consideraciones a la hora de realizar comparaciones entre diferentes efectos. Por

[1] HITCHCOCK, C. *Probabilistic causation*. Stanford: Encyclopedia of Philosophy, 2002.
[2] HUNTER, J. E.; SCHMIDT, F. L. *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*. 2nd Edition.. Newbury Park: Sage Publications, 2004

ejemplo, si los usuarios de un centro deportivo están más satisfechos que los usuarios de otro centro, habiéndose realizado esas mediciones con escalas de diferente rango.

BIBLIOGRAFÍA

COHEN, P.; COHEN, J.; AIKEN, L. S.; WEST, S. G. "The problem of units and the circumstance for POMP". *Multivariate Behavioral Research*. 1999. vol. 34, p. 315-346.

DAWES, J. "Do data characteristics change according to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales". *International Journal of Market Research*, 2008. vol. 50, núm. 1, p. 61-77.

HITCHCOCK, C. *Probabilistic causation*. Stanford: Encyclopedia of Philosophy, 2002.

HUNTER, J. E.; SCHMIDT, F. L. *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*. 2nd Edition. Newbury Park: Sage Publications, 2004.

MARTÍNEZ, J. A. "Estudio de la invarianza de escala mediante el método de cálculo integral en la medición de la calidad percibida de los servicios deportivos". *Revista Internacional de Ciencias del Deporte*. En prensa.

MARTÍNEZ, J. A.; MARTÍNEZ, L. "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*. 2008, vol. 20, núm. 2, p. 305-310.

PARKS, J. B.; SHEWOKIS, P. A.; COSTA, C. A. "Using statistical power analysis in sport management research". *Journal of Sport Management*. 1999, vol. 13, núm. 2, p. 139-147.