

Cita: Lapresa, D.; Otero, A.; Arana, J.; Álvarez, I.; Anguera, M.T. (2021). Concordancia consensuada en metodología observacional: efectos del tamaño del grupo en el tiempo y la calidad del registro. *Cuadernos de Psicología del Deporte*, 21(2), 47-58

Concordancia consensuada en metodología observacional: efectos del tamaño del grupo en el tiempo y la calidad del registro

The consensus agreement method in observational methodology: impact of sample size on time requirements and dataset quality

Concordância consensual na metodologia observacional: efeitos do tamanho do grupo, tempo e qualidade do registro

Lapresa, D.¹, Otero, A.¹, Arana, J.¹, Álvarez, I.² y Anguera, M.T.³

¹Universidad de La Rioja; ²Universidad Isabel I; ³Universidad de Barcelona

RESUMEN

En metodología observacional, para abordar la fiabilidad de los datos ya registrados, suele recurrirse a coeficientes de concordancia, coeficientes de correlación o a la teoría de la generalizabilidad; además, cada vez está tomando mayor protagonismo la concordancia consensuada. Esta forma de concordancia trata de lograr la coincidencia entre los observadores antes del registro. A pesar de su creciente presencia en estudios observacionales, son pocos los trabajos que han profundizado en el desarrollo y optimización de esta forma cualitativa de concordancia. El presente trabajo, además de constituirse en un ejemplo de la utilización de la concordancia por consenso, ha comparado el resultado obtenido (tiempo empleado y ajuste con el registro ideal) por grupos de consenso formados por diferente número de integrantes (dos, tres y cuatro observadores). No se han encontrado diferencias significativas al comparar los grupos de concordancia por consenso de dos, tres y cuatro integrantes, ni en relación al tiempo empleado en el registro, ni en lo relativo al porcentaje de acuerdo con el registro ideal. La determinación del tamaño muestral necesario para obtener diferencias significativas entre los grupos ha permitido elevar conclusiones en términos de eficiencia.

Palabras clave: metodología observacional, concordancia por consenso, tamaño del grupo, tiempo, Kappa de Cohen, ANOVA.

ABSTRACT

The reliability of datasets in observational methodology is typically tested using coefficients of agreement, correlation coefficients, or generalizability theory. Another increasingly popular method used to demonstrate the quality of data is the consensus agreement method, in which two or more observers agree on their coding decisions while creating the dataset. Although the consensus agreement method is being increasingly used in observational studies, few studies have conducted an in-depth analysis of how this qualitative procedure is approached or of how it can be optimized. In this study, in addition to presenting a practical example of the application of the consensus agreement method, we compare the results from three groups (of two, three, and four observers) to analyze performance in terms of time required to code the data and goodness of fit with respect to an optimal dataset. No significant differences were found between the three groups for either of the variables analyzed. Prior calculation of the sample size required to detect significant differences between the groups adds strength to our conclusions regarding the efficiency of the consensus agreement method.

Keywords: observational methodology, consensus agreement, sample size, Cohen's kappa, time, ANOVA.

RESUMIO

Na metodologia observacional, para lidar com a confiabilidade de dois dados já registrados, costumamos passar pelos coeficientes de concordância, coeficientes de correlação ou pela teoria da generalização. Além disso, a concordância consensual vem ganhando cada vez mais destaque. Esta forma de concordância tenta chegar a um acordo entre os observadores antes do registro. Apesar de sua crescente presença em estudos observacionais, poucos estudos se aprofundam no desenvolvimento e otimização de uma forma qualitativa de concordância. Ou apresentar trabalho, além de ser um exemplo do uso de concordância de consenso, comparação ou resultado obtido (tempo despendido e ajuste como registro ideal) por grupos de consenso formados por diferentes números de membros (dois, três e quatro observadores). Não foram encontradas diferenças significativas na comparação de dois grupos de concordância por consenso de dois, três e quatro membros, não em relação ao tempo gasto não registrado, mas em relação ao percentual de concordância conforme lista ideal. A determinação do tamanho dá a amostra necessária para obter diferenças significativas entre os grupos permitindo-nos tirar conclusões em termos de eficiência.

Palavras chave: metodologia observacional, concordância consensual, tamanho da amostra, Kappa de Cohen, tempo, ANOVA.

INTRODUCCIÓN

De acuerdo con Anguera (1990) existen tres formas diferentes que permiten contemplar la fiabilidad de los datos observacionales. En primer lugar, la concordancia entre observaciones: una medida -la más utilizada es el coeficiente Kappa de Cohen (1960)- en la que un observador en varias ocasiones o varios observadores, trabajando de forma independiente y autónoma, concuerdan en el registro de las conductas a partir de un mismo instrumento de observación. En segundo lugar, la teoría psicométrica de la fiabilidad; a partir de las definiciones de fiabilidad de la teoría clásica psicométrica, se recurre a un coeficiente de correlación. En tercer lugar, en el seno de la teoría de la Generalizabilidad se consideran todas las fuentes de variación que afectan a los resultados observados, integrándolas en una estructura global (Cronbach, Gleser, Nanda y Rajaratnam, 1972). Además, Anguera (1990) añade la concordancia por consenso, que es la que se constituye en objeto de estudio en el presente trabajo.

Esta concordancia -que cobra, día tras día, mayor protagonismo en metodología observacional- trata de lograr la coincidencia interpretativa entre los observadores antes del registro. Entre las ventajas que posee la concordancia por consenso, destaca la obtención de un registro único, además de que el instrumento de observación se ve fortalecido al quedar mejor perfilados tanto sus definiciones como los matices que deban añadirse. En la otra vertiente -es decir, la de los inconvenientes-, destacar las posibles disfunciones en el trabajo de los grupos por consenso: enfrentamientos que puedan dar lugar a dificultades para lograr el consenso entre los observadores; o, por el contrario, que en el grupo pese el prestigio o la elevada competencia de uno de los observadores (Vigil-Colet, Morales-Vives y Lorenzo-Seva, 2013).

Numerosos son los estudios que utilizan esta forma cualitativa para llevar a cabo el control de la calidad del dato en sus investigaciones -Castellano, Hernández-Mendo,

Gómez de Segura, Fontetxa y Bueno (2000), Castellano y Hernández-Mendo (2003), Gamonal, Muñoz, León, e Ibáñez (2018), Garay, Hernández-Mendo y Morales (2006), Gorospe, Hernández-Mendo, Anguera y Martínez de Santos (2005), Castellano, Perea y Hernández-Mendo (2008), Jurado, Anguera, Reigal, y Hernández-Mendo (2021), Montesano, Feixas y Varlotta (2009), Hernández-Mendo, Díaz y Morales (2010), Hernández-Mendo y Planchuelo (2012), Usabiaga, Castellano, Blanco-Villaseñor y Casamichana (2013), Ardá, Maneiro, Rial, Losada y Casal (2014), Morillo-Baro, Reigal y Hernández-Mendo (2015), Arana, Lapresa, Anguera y Garzón (2016)-, pero pocos son los que profundizan en la forma en la que se ha desarrollado la concordancia consensuada.

Es por ello, que el presente trabajo pretende constituirse en un ejemplo concreto de la utilización de este tipo de concordancia cualitativa, así como satisfacer el objetivo de analizar el efecto del tamaño del grupo (dos, tres y cuatro integrantes) en la concordancia por consenso, en lo relativo a: el tiempo empleado en el registro y la calidad del registro (concordancia con el registro ideal).

MATERIAL Y MÉTODOS

Participantes

La selección de los participantes fue censal, incluyendo a los alumnos que han cursado la asignatura "Observación Sistemática en la Escuela Infantil" correspondiente al segundo curso del Grado de Educación Infantil de la Universidad de La Rioja.

El Gran Grupo de esta asignatura, estaba compuesto 76 alumnos. Se aplicó, como criterio de exclusión, la no asistencia a las sesiones que dieron soporte al presente trabajo, que fueron nueve alumnos. Por tanto, el número de participantes se limitó a 67.

Los integrantes del registro mediante concordancia consensuada se distribuyeron en tres Grupos Reducidos,

Concordancia consensuada en metodología observacional

dando lugar a veintitrés subgrupos de concordancia por consenso distribuidos al azar -en el seno de cada Grupo Reducido mediante la función “aleatorio” de la hoja de cálculo *Microsoft Office Excel*, 2016-, compuestos por dos, tres o cuatro integrantes (véase tabla 1).

También han intervenido en el desarrollo de este estudio, dos observadores -autores del presente trabajo- con formación en metodología observacional, a los que más adelante se les denominará Observador 1 y Observador 2.

Instrumentos

El primer paso fue decidir la sesión de registro y el instrumento de observación que permitiera satisfacer los objetivos del estudio. En este sentido se parte de la consideración previa de que se pretende: a) provocar que el consenso entre participantes no sea evidente; b) que la herramienta observacional sea sencilla en cuanto a su comprensión y utilización.

De esta forma, para la elección de la sesión de observación (un partido de fútbol-3) se ha tenido en consideración la variabilidad del registro de la dimensión zona en estudios observacionales. De acuerdo con O'Donoghue (2007), el criterio zona es común fuente de variación entre registros; es decir, es una dimensión que en lo relativo a la fiabilidad del dato suele presentar discrepancias entre observaciones. Además, las dificultades del niño en la iniciación más temprana en su relación con el balón, hace que en el fútbol practicado por estos niños proliferen los intentos de control (Lapresa, Arana y Garzón, 2020) de lo que se deriva la dificultad de discriminar si existe o no contacto. Es por ello que se seleccionó como sesión de observación un partido de fútbol-3 practicado por niños de categoría prebenjamín (primer curso de Educación Primaria); concretamente la fase ofensiva del equipo vestido de rojo.

El instrumento de observación utilizado es una combinación de formato de campo y dos sistemas de categorías, exhaustivos y mutuamente excluyentes (Anguera, 2003). Dos son los criterios vertebradores del instrumento de observación: Zona de inicio del contacto - la zona en la que se encuentra el balón cuando el jugador contacta con él, por primera vez- $ZI=\{ZI10, ZI20, ZI30, ZI40, ZI50, ZI60, ZI41, ZI51, ZI61, ZI70, ZI80, ZI90\}$; Zona de final del contacto -la zona donde el jugador contacta por última vez con el balón- $ZF=\{ZF10, ZF20, ZF30, ZF40, ZF50, ZF60, ZF41, ZF51, ZF61, ZF70, ZF80, ZF90\}$. Se ha tomado como referencia la distribución espacial del terreno de juego, en 12 zonas diferentes, para fútbol-3 de Lapresa, Arana, Anguera, Pérez y Amatria (2016).

Procedimiento

El procedimiento seguido para conseguir y avalar los datos que soportan los análisis que permiten elevar las

conclusiones que satisfacen el objetivo prefijado en el presente trabajo, se estructura en cuatro fases: a) la determinación del registro ideal; b) la fiabilidad del proceso de alineación; c) la formación y el registro de los grupos de concordancia por consenso; d) el cálculo de la concordancia entre los registros de los grupos de concordancia por consenso y el registro ideal.

a) Determinación del registro ideal

Para la determinación del registro ideal se ha recurrido a dos observadores (observadores 1 y 2), con una larga trayectoria en la realización de estudios observacionales. Se considerará que nos encontramos ante el registro ideal si los registros efectuados por ambos observadores coinciden totalmente (se obtiene, bajo el parámetro orden, un porcentaje de acuerdo del 100% y un Kappa de Cohen= 1). Mediante el *software* Lince, versión 1.2.1 (Gabin et al., 2012) ambos participantes han registrado la misma sesión de observación que los grupos por consenso. Debe destacarse que han contado con unas ventajas de las que no han podido disponer los grupos por consenso: a) la posibilidad de ampliar el tamaño de la imagen mediante un reproductor de video (VLC) externo al *software* Lince; b) la ausencia de requerimiento temporal; que ha permitido que estos observadores se tomasen todo el tiempo que necesitasen para realizar el registro.

b) De la fiabilidad del proceso de alineación

Del proceso de alineación (con el registro ideal) de la totalidad de los registros de los grupos de concordancia por consenso se ha encargado el Observador 1. Con el fin de garantizar la fiabilidad de los resultados, el segundo observador ha realizado el proceso de alineación de cuatro registros, lo que constituye un 18% de los grupos de concordancia por consenso; un porcentaje lo suficientemente amplio para avalar la fiabilidad del proceso de alineación, que además supone una economía de esfuerzo de cara al investigador, y teniendo en cuenta que en la literatura científica de diversos ámbitos hay numerosos trabajos que para hallar la concordancia utilizan un porcentaje de registros comprendido entre el 5 y el 20% (Alcover et al., 2019; Alsasua, Lapresa, Arana y Anguera, 2019; Amatria, Lapresa, Arana y Anguera, 2016; Belasko, Herrán y Anguera, 2019; Belza, Herrán y Anguera, 2020; Diana et al., 2017; Lapresa, Arana, Amatria, Fernández y Anguera, 2017; Lapresa et al., 2016; Planes y Anguera, 2015; Sagastui, Herrán y Anguera, 2020; Sastre, Lapresa, Arana, Ibáñez y Anguera, 2021), si bien en algunos trabajos, de forma minoritaria, se halla la concordancia a partir de la totalidad del registro (Vázquez-Diz et al., 2019a, 2019b).

El procedimiento de alineación se ha llevado a cabo mediante *Microsoft Office Excel 2016*, teniendo en consideración las siguientes premisas de cara a la interpretación de la alineación de los multieventos: a) ajuste

en base al tiempo del registro; b) ajuste en cuanto las zonas (de inicio y finalización) que constituyen los multieventos.

En el caso de que no hubiese correspondencia entre registros a partir de estas premisas, se deja una fila del registro correspondiente -ya sea el del grupo por consenso o el ideal- en blanco (a excepción del punto requerido por la sintaxis de codificación de GSEQ5 en datos multievento). De esta forma, se va alineando tanto el registro ideal como, en todos los grupos, el registro por consenso (véase figura 1). Y es que, para garantizar la fiabilidad del proceso de alineación, será necesario la comparación de la alineación efectuada por el Observador 1 y Observador 2, tanto del registro ideal como del registro del grupo encargado de la concordancia por consenso.

c) Grupos de concordancia por consenso (formación y registro)

Para hacer posible este trabajo se organizaron con los alumnos del grado de Educación Infantil de la asignatura “Observación sistemática en la escuela infantil”, dos sesiones de trabajo denominadas “Formación” y “Registro por consenso”. Las dos sesiones planteadas, de una hora cada una, se realizaron a finales de marzo del año 2017 durante dos semanas consecutivas, en el horario correspondiente a cada Grupo Reducido de la asignatura.

- Primera sesión: formación de los observadores:

En esta primera jornada, los alumnos trabajaron a nivel individual cada uno con su ordenador. En primer lugar, se explicó detalladamente el instrumento de observación. Posteriormente, cada alumno, descargó el programa Lince, versión 1.2.1 (Gabin et al., 2012), e introdujo el instrumento de observación en el seno del programa de registro y codificación. Para facilitar el proceso de enseñanza-aprendizaje se preparó una hoja de instrucciones específica para esta primera sesión, con los contenidos y pautas necesarias para la comprensión de la herramienta observacional y del instrumento de registro.

Por último, se destinaron los últimos 10 minutos de la sesión de formación, para que los alumnos registrasen cinco minutos del mismo partido, pero distintos a los que se iban a utilizar en el registro mediante concordancia consensuada.

- Segunda sesión: registro mediante concordancia consensuada:

Antes de comenzar la sesión de registro se procedió a la instalación en los ordenadores requeridos del *software* de registro. Además, se añadió al escritorio una carpeta con el archivo “ilince” (instrumento de observación) y los cinco minutos de sesión por registrar.

Se fue llamando a cada integrante y ubicándolo en los diferentes puestos de registro. Cuando algún alumno no asistió a la sesión, los grupos por consenso se “convirtieron” en grupos del número de participantes presentes.

Antes de comenzar el registro por consenso, se realizó un repaso del instrumento de observación y de la utilización del programa y se les otorgó una segunda hoja de instrucciones. Posteriormente, a partir de Arana, et al. (2016), se explicaron las pautas para el funcionamiento del grupo por consenso: “Enfocad la tarea en base a la lógica; no cambiéis de opinión por evitar el conflicto; apoyad las ideas con las que en parte estéis en concordancia (sinergia); considerad las diferencias de opinión como algo positivo (no penséis en términos de ganar-perder); defended opiniones con argumentos; evitad el procedimiento: “ahora cedo yo y luego tú”; escuchad activamente: “tenemos que escuchar y ser escuchados”; no critiquéis; aportad sugerencias; todos tenéis la misma responsabilidad en cuidar el proceso”.

Antes de dar comienzo al registro por consenso, se proyectó en pantalla un cronómetro, con la indicación de que los participantes anotasen el tiempo empleado en la realización del registro; además, justo al acabar, debían levantar la mano y llamar la atención al respecto. Posteriormente, se procedió a la exportación de los resultados en formato Excel y GSEQ5 -datos de multievento (Bakeman y Quera, 1995)-.

d) Cálculo de la concordancia entre el registro de los grupos de concordancia por consenso y el registro ideal.

Tras haberse garantizado la fiabilidad del proceso de alineación, el Observador 1 ha procedido a la alineación de las conductas registradas por cada grupo de concordancia por consenso y, en correspondencia a dicho registro, las del registro ideal (véase figura 1).

Posteriormente, se ha procedido a calcular cuantitativamente la concordancia entre observaciones en la determinación de la concordancia entre los grupos de concordancia por consenso y el registro ideal. A través del programa informático GSEQ5 y teniendo en consideración los trabajos de (Bakeman y Quera, 1995, 2011), se ha procedido al cálculo del porcentaje de acuerdo y del coeficiente Kappa de Cohen entre observaciones. Al respecto, GSEQ5 tiene una limitación en este cálculo con datos de multievento -al incluir en el cálculo, tan solo, el primer código del multievento-. Esta limitación ha sido superada (Garzón, 2010), gracias a las posibilidades que otorga la sintaxis de codificación de GSEQ 5, uniendo los eventos constitutivos de cada uno de los multieventos con un guión bajo; de esta forma, cualquier diferencia en el registro de cada una de las dos dimensiones que componen el registro de cada contacto se trata como discrepancia en el cálculo del porcentaje de acuerdo y del coeficiente Kappa de Cohen.

Concordancia consensuada en metodología observacional

Tabla 1. Distribución de grupos de concordancia por consenso en cada Grupo Reducido.

Grupo Reducido	2 observadores	3 observadores	4 observadores
GR-1	3 grupos	3 grupos	2 grupos
GR-2	2 grupos	3 grupos	2 grupos
GR-3	3 grupos	3 grupos	2 grupos

Observador 1						Observador 2					
Registro grupo consenso			Registro ideal			Registro grupo consenso			Registro ideal		
Tsegundos	zona inicio	zona final	Tsegundos	zona inicio	zona final	Tsegundos	zona inicio	zona final	Tsegundos	zona inicio	zona final
00:32,8	z141	zf40	00:32,8	z150	zf50	00:32,8	z141	zf40	00:32,8	z150	zf50
00:35,9	z150	zf51	00:35,8	z150	zf50	00:35,9	z150	zf51	00:35,8	z150	zf50
00:45,8	z160	zf60	.	.	.	00:45,8	z160	zf60	.	.	.
01:20,8	z150	zf50	00:48,2	z161	zf90	.	.	.	00:48,2	z161	zf90
01:25,1	z151	zf51	01:21,0	z120	zf30	01:20,8	z150	zf50	01:21,0	z120	zf30
.	.	.	01:24,9	z151	zf51	01:25,1	z151	zf51	01:24,9	z151	zf51
02:11,1	z130	zf51	02:00,9	z120	zf20	.	.	.	02:00,9	z120	zf20
02:16,1	z151	zf50	02:10,8	z160	zf50	02:11,1	z130	zf51	02:10,8	z160	zf50
02:27,2	z180	zf80	02:14,6	z150	zf50	02:16,1	z151	zf50	02:14,6	z150	zf50
02:46,1	z190	zf90	02:27,9	z151	zf80	02:27,2	z180	zf80	02:27,9	z151	zf80
.	.	.	02:45,9	z190	zf90	02:46,1	z190	zf90	02:45,9	z190	zf90
02:52,9	z180	zf70	02:48,2	z190	zf80	.	.	.	02:48,2	z190	zf80
03:01,1	z141	zf40	02:54,6	z170	zf70	02:52,9	z180	zf70	02:54,6	z170	zf70
03:22,6	z110	zf10	.	.	.	03:01,1	z141	zf40	.	.	.
04:04,9	z140	zf20	03:22,9	z110	zf10	03:22,6	z110	zf10	03:22,9	z110	zf10
05:02,1	z190	zf80	04:03,9	z140	zf40	04:04,9	z140	zf20	04:03,9	z140	zf40
.	.	.	04:59,7	z161	zf61	05:02,1	z190	zf80	04:59,7	z161	zf61

Figura 1. Alineación del registro ideal con el registro de un grupo por consenso, realizado por observador 1 y observador 2.



Figura 2. Captura del instrumento de observación en el seno del programa de registro y codificación Lince.

Registro y codificación de los datos

En el presente trabajo, el registro mediante el programa informático Lince versión 1.2.1 (véase figura 2) se ha llevado a cabo a partir del video facilitado, y ha consistido en la observación de la fase ofensiva del equipo rojo. La sesión de observación, de cinco minutos de duración, está compuesta por un determinado número de contactos -que tienen una zona de inicio y una zona de finalización-.

Se considera contacto cuando un jugador del equipo rojo realiza, al menos, dos toques sobre el balón. El contacto se inicia con el primer toque al balón (criterio zona de inicio) y acaba con el último toque al balón (criterio zona final). Los contactos que se desarrollan entre estos dos toques -zonas de desarrollo- no se registran. Un toque del jugador rival o una interrupción reglamentaria da lugar al fin del contacto. Se registra la posición en la que se encuentra el jugador cuando ha realizado el contacto con el balón, no el lugar en el que acaba el balón. Destacar que el campo donde se juega el partido de fútbol está dividido a través de conos y piezas planas de goma para facilitar el entendimiento del reparto zonal efectuado. Los saques de banda o de esquina no se consideran contacto. Los saques del portero con la mano no se consideran contacto; si el portero realiza un auto-saque (deposita el balón con la mano en el suelo y comienza a jugar con el pie) el primer contacto con el pie supone el inicio del contacto.

De acuerdo con la clasificación de Bakeman (1978) los datos registrados han sido tipo IV, al ser concurrentes (incorporan las dos dimensiones del instrumento de observación) y tiempo-base. No obstante, los análisis realizados en el presente trabajo se han realizado bajo el parámetro orden -transformándose en datos tipo II; concurrentes y evento-base (Bakeman y Quera, 1995)-.

RESULTADOS

Del registro ideal

Podemos afirmar que se ha obtenido el registro ideal, ya que tanto el Observador 1 como el Observador 2, ambos expertos, han obtenido el mismo registro -porcentaje de acuerdo del 100% y Kappa de Cohen= 1- de la sesión de observación que posteriormente han registrado los grupos de concordancia por consenso.

Resultados de la fiabilidad del proceso de alineación

En lo relativo a la fiabilidad del proceso de alineación, mediante la comparación de la alineación efectuada por el segundo observador de cuatro registros de grupos de concordancia por consenso -lo que supone la alineación interrelacionada tanto del registro ideal como del registro del grupo encargado de la concordancia por consenso (véase figura 1)-, se ha obtenido un porcentaje de

acuerdo del 100% y un valor de Kappa de Cohen= 1. De esta forma, queda garantizada la fiabilidad del proceso de ordenación.

Tiempo empleado en el registro y concordancia cuantitativa entre los grupos de concordancia por consenso y el registro ideal

En el presente trabajo de investigación se ha garantizado la constancia intersesional mediante el cumplimiento de los siguientes requisitos: a) todos los registros se efectuaron en el aula y equipos referidos; ninguna sesión de Grupo Reducido fue interrumpida; todo grupo de concordancia por consenso anotó el tiempo empleado en su registro nada más finalizado; ninguno de los grupos de concordancia por consenso analizados pudo tener menos de la mitad o más del doble de los multieventos constitutivos del registro ideal. De esta forma, se evita incluir en el análisis registros correspondientes a grupos que pudieron presentar serias deficiencias en su funcionamiento. En la tabla 2 se expone, del registro correspondiente a cada grupo de concordancia por consenso: el número de multieventos constitutivo del registro, el tiempo empleado en el registro, el porcentaje de acuerdo y el Kappa de Cohen entre los grupos de concordancia por consenso y el registro ideal.

Tiempo empleado en el registro en función del tamaño del grupo por consenso

Para contrastar si el tiempo empleado en la observación (en segundos) es significativamente diferente en función del número de observadores que integran los grupos, se realiza un Análisis de Varianza (ANOVA), ya que se está trabajando con datos cuantitativos (véase tabla 3). La significación encontrada (0,955) es superior a 0,05 por lo que no se puede rechazar la hipótesis de igualdad de medias.

Acuerdo entre observaciones en función del tamaño del grupo por consenso

El rango promedio de porcentaje de acuerdo con el registro ideal, es superior en los grupos de cuatro componentes (15,25), seguido por los de tres y dos participantes. Esto indica que el grado de acuerdo con el registro ideal de las observaciones realizadas por los grupos de cuatro es superior al alcanzado en los grupos de tres (10,61) y dos componentes (9,43). Para comprobar si el porcentaje de acuerdo encontrado entre los diferentes evaluadores es significativamente distinto en función del número de observadores que componen los grupos, y debido a que se está trabajando con porcentajes, se ha llevado a cabo la prueba no paramétrica de Kruskal-Wallis (Breslow, 1970), no habiéndose obtenido diferencias significativas ($\chi^2= 2,907$; g.l.= 2; p= 0,234).

Concordancia consensuada en metodología observacional

Tabla 2. Número de multieventos, tiempo del registro, porcentaje de acuerdo y Kappa de Cohen entre los grupos de concordancia por consenso y el registro ideal. En la columna grupo, el primer número es el de integrantes y el segundo el identificador. El grupo 2.4 no cumple los requisitos de constancia intersesional.

Grupo	Registros	Tiempo	%Acuerdo	Kappa de Cohen
2-1	18	7'28''	25%	0,19
2-2	20	12'30"	20%	0,11
2-3	16	12'50''	33%	0,29
2-4	7	10'35''	No requisitos ci	No requisitos ci
2-5	14	13'20''	12%	0,08
2-6	15	19'30''	31%	0,28
2-7	14	15'24''	31%	0,28
2-8	12	22'00''	44%	0,39
3-1	16	18'40''	35%	0,35
3-2	14	10'01''	27%	0,24
3-3	13	16'04''	40%	0,37
3-4	9	9'02''	6%	0,02
3-5	14	19'06''	19%	0,16
3-6	13	17'28''	44%	0,40
3-7	13	10'15''	27%	0,24
3-8	15	9'21''	38%	0,35
3-9	15	19'47''	18%	0,11
4-1	18	9'00''	33%	0,31
4-2	14	18'36''	53%	0,50
4-3	14	14'30''	31%	0,26
4-4	12	13'50''	31%	0,28
4-5	16	16'50''	39%	0,36
4-6	18	18'51''	39%	0,35

Tabla 3. Tiempo empleado en el registro en función del tamaño del grupo por consenso: fuentes de variabilidad.

	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Componentes de los grupos	6644,739	2	3322,369	0,046	0,955
Error	1378140,579	19	72533,715		
Total	1384785,318	21			

Estimación del tamaño muestral necesario para encontrar diferencias significativas

El tamaño muestral indica el número de elementos contenidos en la muestra analizada. Su determinación se realiza en base a dos criterios: un tamaño muestral elevado proporciona análisis estadísticos más precisos; y un tamaño muestral reducido conlleva una mayor economía en el desarrollo del estudio. Lo ideal es conjugar ambas premisas para determinar un tamaño muestral lo suficientemente grande para que las conclusiones se consideren fiables, pero suficientemente pequeño para no tener que recoger un excesivo número de datos que no contribuyan a la calidad del diseño.

Para estimar el tamaño idóneo de la muestra (Lapresa, Álvarez, Anguera, Arana y Garzón, 2015), se ha utilizado la versión 3.1.2 del *software* G-Power (véase figura 3), a partir de Faul, Erdfelder, Lang y Buchner (2007) y Faul, Erdfelder, Buchner y Lang (2009). Este software permite introducir determinados datos a priori para calcular el tamaño muestral necesario. Si el tamaño muestral estimado es razonable se puede optar por replicar el estudio con una nueva muestra más amplia.

En concreto, en el presente trabajo, teniendo en cuenta que se trabaja con tres grupos (de 2, de 3 y de 4 observadores) y a partir de las medias cuadráticas obtenidas en el ANOVA (Media Cuadrática Inter-grupos: 3322,369; Media Cuadrática Intra-grupos: 72533,715) se ha obtenido un tamaño del efecto de 0,214 -que conforme los valores de

referencia convencionalmente aceptados (Cárdenas y Arancibia, 2016) puede considerarse un tamaño del efecto medio-. Posteriormente se ha seleccionado una probabilidad de cometer error tipo I de 0,05 y una potencia estadística de 0,95, obteniéndose un tamaño de la muestra de 342 elementos.

Repartidos entre los tres tipos de grupos de concordancia por consenso (de dos, tres y cuatro componentes), cada uno debería contener 114 puntuaciones (grupos). Así, el número de participantes necesario para completar el estudio conforme la simulación sería de 1026 observadores.

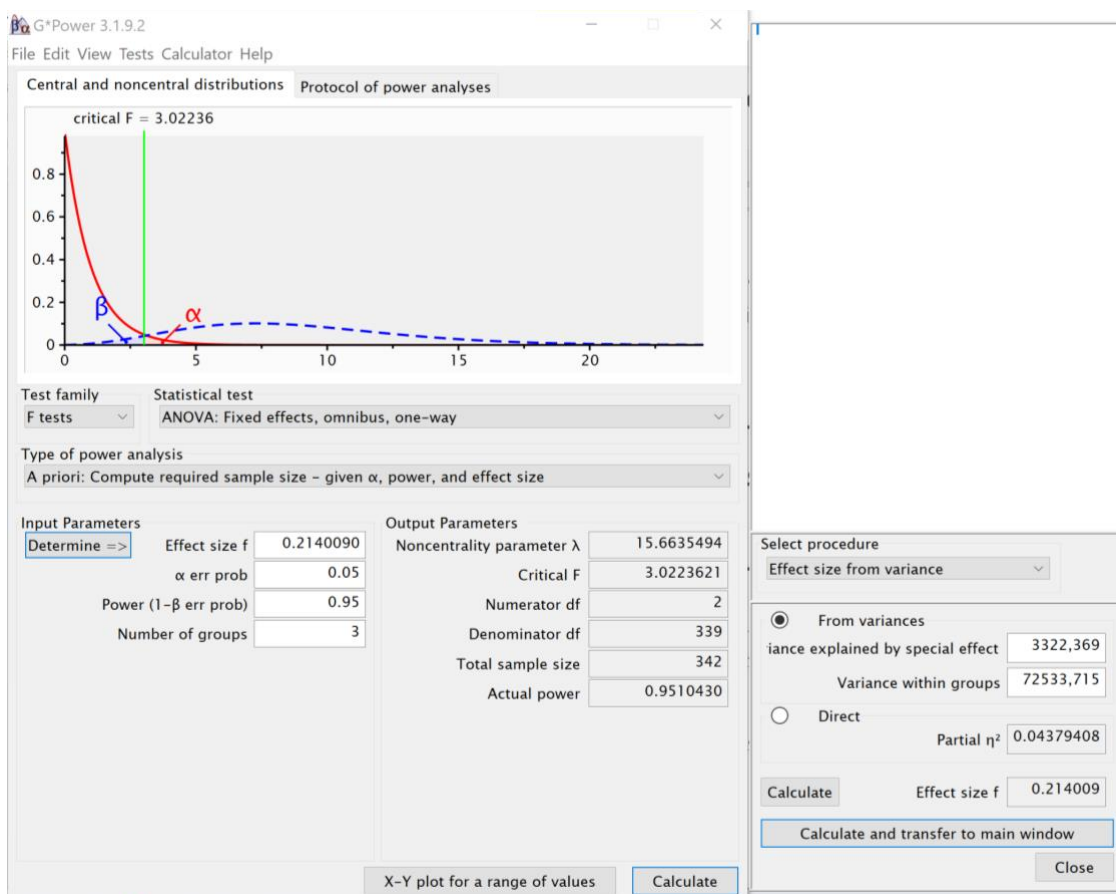


Figura 3. Estimación del tamaño de la muestra, a partir de las medias cuadráticas obtenidas en el ANOVA, mediante el *software G-Power*, versión 3.1.2.

DISCUSIÓN

A pesar de la creciente presencia de la concordancia consensuada en estudios observacionales, son pocos los trabajos que han profundizado en el desarrollo y optimización de esta forma cualitativa de concordancia. El presente trabajo, se ha constituido en un ejemplo concreto de utilización de este tipo de concordancia cualitativa (Anguera, 1990), que puede servir de referencia para futuros trabajos.

A continuación, se presentan las conclusiones de este trabajo que pretende analizar el efecto del tamaño del grupo (dos, tres y cuatro integrantes) en la concordancia por

consenso, en lo relativo al tiempo empleado y a la calidad del registro (concordancia con el registro ideal).

Al respecto, no se han encontrado diferencias significativas al comparar los grupos de concordancia por consenso número de dos, tres y cuatro integrantes: a) ni en relación al tiempo empleado en el registro; b) ni en lo relativo al porcentaje de acuerdo con el registro ideal.

Para que dichas diferencias resultasen significativas se debería disponer de, como mínimo, 1026 observadores. Dado que se trataría de una muestra muy difícil de gestionar por su amplitud, no consideramos adecuado replicar este trabajo en circunstancias similares.

Concordancia consensuada en metodología observacional

No obstante, sí que consideramos recomendable replicar este estudio en otras condiciones:

a) optimizando la motivación de los observadores (como, por ejemplo, mediante la obtención de un porcentaje de la nota como consecuencia del ajuste al registro ideal). Así lo recomiendan los reducidos valores de porcentaje de acuerdo y Kappa de Cohen obtenidos (Landis y Koch, 1977) en relación al registro ideal, así como que la confiabilidad de un observador puede estar limitada por efectos biosociales, biográficos, psicosociales, situacionales o de expectativa (Anguera, Blanco, Losada y Sánchez-Algarra, 1999; Losada y Manolov, 2015).

b) optimizando el proceso de formación de los observadores, ya que ha quedado demostrado que la competencia adquirida por el observador está condicionada por el entrenamiento que haya recibido (Arias, Argudo y Alonso, 2009; Losada y Manolov, 2015; Usabiaga y Castellano, 2014).

Al no haberse encontrado en el presente trabajo ninguna diferencia significativa entre los resultados obtenidos entre los grupos de concordancia por consenso de dos, tres y cuatro integrantes, lo razonable sería optar por grupos pequeños, dada la economía que resulta de su empleo. Ahora bien, teniendo en cuenta que la concordancia elevada entre dos observadores -ambos podrían coincidir en el error del registro- no implica precisión (Blanco-Villaseñor y Anguera, 2003; Mitchell, 1979), la solución más operativa sería la utilizada en la mayoría de los trabajos que utilizan concordancia por consenso de recurrir a tres observadores (Arana, et al., 2016; Castellano, et al., 2000; Gorospe, et al., 2005; Hernández-Mendo et al., 2010; Hernández-Mendo y Planchuelo, 2012). Aunque, de esta forma, *mutatis mutandis*, se podría justificar perfectamente la propuesta de si 'cuatro mejor que tres' utilizada en Montesano, et al. (2009).

Además, en la conformación de los grupos de concordancia por consenso se recomienda optimizar la formación de los observadores (Anguera, 2003) y la conformación de grupos (criterio de homogeneidad que evite las relaciones verticales *versus* criterio de heterogeneidad en relación a la formación relativa al objeto de estudio) con el objetivo de equilibrar el prestigio o dominio (igualdad de estatus) de los integrantes del grupo (Arana, et al., 2016).

APLICACIONES PRÁCTICAS

El presente trabajo pretende servir de referencia para investigadores que estén pensando acudir a la concordancia consensuada para garantizar la fiabilidad de sus registros. Además del desarrollo teórico se ofrecen pautas operativas para optimizar la organización del proceso de registro utilizando este tipo de concordancia cualitativa. En lo relativo al número de observadores, la

solución más operativa sería la que cuenta con mayor presencia en estudios observacionales que utilizan este tipo de concordancia, recurriéndose a tres observadores; quedando también justificada, a partir de los resultados del presente trabajo, la propuesta de cuatro observadores.

Agradecimientos

Los autores agradecen el apoyo del subproyecto *Vías de integración entre datos cualitativos y cuantitativos, desarrollo del caso múltiple, y synthesis review como ejes principales para un futuro innovador en investigación de actividad física y deporte* [PGC2018-098742-B-C31] (2019-2021) (Ministerio de Ciencia, Innovación y Universidades / Agencia Estatal de Investigación / Fondo Europeo de Desarrollo Regional), que forma parte del proyecto coordinado *New approach of research in physical activity and sport from mixed methods perspective* (NARPAS_MM) [SPGC201800X098742CV0]; así como del proyecto *Tecnología i aplicació multimedia i digital als dissenys observacionals* [2014 SGR 971], Generalitat de Catalunya Research Group, Grup de recerca i innovació en dissenys (GRID). Además, este estudio ha contado con financiación de la Universidad de La Rioja.

REFERENCIAS (APA 7ª EDICION)

1. Alcover, C, Mairena, M.A., Mezzatesta, M., Elías, N., Díez, M., Balañá, G., ... y Arias-Pujol, E. (2019). Mixed methods approach to describe social interaction during a group intervention for adolescents with Autism Spectrum Disorders. *Frontiers in Psychology*, 10:1158. <http://dx.doi.org/10.3389/fpsyg.2019.01158>
2. Alsasua, R., Lapresa, D., Arana, J., y Anguera, M.T. (2019). A log-linear analysis of efficiency in elite basketball applied to observational methodology. *International Journal of Sports Science & Coaching*, 14(3), 363-371. <https://doi.org/10.1177/1747954119837819>
3. Amatria, M., Lapresa, D., Arana, J., Anguera, M.T., y Garzón, B. (2016). Optimization of Game Formats in U-10 Soccer Using Logistic Regression Analysis. *Journal of Human Kinetics*, 54, 163-171.
4. Anguera, M.T. (1990). Metodología observacional. En J. Arnau, M.T. Anguera, y J.

- Gómez (Eds.), *Metodología de la investigación en Ciencias del Comportamiento* (pp. 125-238). Murcia: Universidad de Murcia.
5. Anguera, M.T. (2003). La observación. En C. Moreno Rosset (Ed.), *Evaluación psicológica. concepto, proceso y aplicación en las áreas del desarrollo y de la inteligencia* (pp. 271-308). Madrid: Sanz y Torres.
 6. Anguera, M.T., Blanco-Villaseñor, A., Losada, J.L., y Sánchez-Algarra, P. (1999). Análisis de la competencia en la selección de observadores. *Metodología de las Ciencias del Comportamiento*, 1(1), 95-114.
 7. Arana, J., Lapresa, D., Anguera, M.T., y Garzón, B. (2016). *Ad hoc* procedure for optimising agreement between observational records. *Anales de Psicología*, 32(2), 589-595. <http://dx.doi.org/10.6018/analesps.32.2.213551>
 8. Ardá, T., Maneiro, R., Rial, A., Losada, J.L., y Casal, A. (2014). Análisis de la eficacia de los saques de esquina en la copa del mundo de fútbol 2010. Un intento de identificación de variables explicativas. *Revista de Psicología del Deporte*, 23(1), 165-172.
 9. Arias, J.L., Argudo, F., y Alonso, J.I. (2009). El proceso de formación de observadores y la obtención de la fiabilidad en metodología observacional para analizar la dinámica de juego en minibásquet. *Revista Apunts. Educación Física y Deportes*, 4, 40-45.
 10. Belasko, M., Herrán, E., y Anguera, M.T. (2019). Dressing toddlers at the Emmi Pikler nursery school in Budapest: caregiver instrumental behavioral pattern. *European Early Childhood Education Research Journal*, 27(6), 972-887. <https://doi.org/10.1080/1350293X.2019.1678928>
 11. Bakeman, R. (1978). Untangling streams of behavior: Sequential analysis of observation data. In G.P. Sackett (Ed.) *Observing Behavior, Vol. 2: Data collection and analysis methods*, 63-78. Baltimore: University of Park Press.
 12. Bakeman, R. y Quera, V. (1995). *Analyzing interaction: Sequential analysis with SDIS and GSEQ*. Cambridge: Ed. Cambridge University Press.
 13. Bakeman, R. y Quera, V. (2011). *Sequential Analysis and Observational Methods for the Behavioral Sciences*. Cambridge: Cambridge University Press.
 14. Belza, H., Herran, E., y Anguera, M.T. (2020). Early childhood, breakfast and related tools: Analysis of adults' function as mediators. *European Journal of Psychology of Education*, 25, 495-527. <https://doi.org/10.1007/s10212-019-00438-4>
 15. Blanco-Villaseñor, A. y Anguera, M.T. (2003). Calidad de los datos registrados en el ámbito deportivo. En A. Hernández-Mendo (Coord.), *Psicología del Deporte* (Vol. 2). *Metodología* (pp. 35-73). Buenos Aires: Efdportes (www.efdeportes.com).
 16. Breslow, N. (1970). A generalized Kruskal-Wallis test for comparing K samples subject to unequal patterns of censorship. *Biometrika*, 57(3), 579-594. <https://doi.org/10.2307/2334776>
 17. Cárdenas, M. y Arancibia, H. (2016). Potencia estadística y cálculo del tamaño del efecto en G*Power: complementos a las pruebas de significación estadística y su aplicación en psicología. *Salud & Sociedad*, 5(2), 210-244. <https://doi.org/10.22199/S07187475.2014.0002.0006>
 18. Castellano, J., Hernández-Mendo, A., Gómez de Segura, P., Fontetxa, E., y Bueno, I. (2000). Sistema de codificación y análisis de la calidad del dato en el fútbol de rendimiento. *Psicothema*, 12(4), 635-641.
 19. Castellano, J., Perea, A., y Hernández-Mendo, A. (2008). Análisis de la evolución del fútbol a lo largo de los mundiales. *Psicothema*, 20(4), 928-932.
 20. Castellano, J. y Hernández-Mendo, A. (2003). El análisis de coordenadas polares para la estimación de relaciones en la interacción motriz en fútbol. *Psicothema*, 15(4), 569-574.
 21. Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37-46. <https://doi.org/10.1177/001316446002000104>
 22. Cronbach, L.J., Gleser, G.C., Nanda, H., y Rajaratnam, N. (1972). *The dependability of*

Concordancia consensuada en metodología observacional

- behavioral measurements: theory of generalizability for scores and profiles*. New York: Wiley.
23. Diana, B., Zurloni, V., Elia, M., Cavallera, C.M., Jonsson, G.K., y Anguera, M.T. (2017). How game location affects soccer performance: T-pattern analysis of attack actions. *Frontiers in Psychology*, 8:1415. <https://doi.org/10.3389/fpsyg.2017.01415>
 24. Faul, F., Erdfelder, E., Buchner, A., y Lang, A.G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41, 1149-1160. <https://doi.org/10.3758/BRM.41.4.1149>
 25. Faul, F., Erdfelder, E., Lang, A.G., y Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175-191. <https://doi.org/10.3758/BF03193146>
 26. Gabin, B., Camerino, O., Anguera, M.T., y Castañer, M. (2012). Lince: multiplatform sport analysis software. *Procedia-Social and Behavioral Sciences*, 46, 4692-4694. <https://doi.org/10.1016/j.sbspro.2012.06.320>
 27. Gamonales, J.M., Muñoz, J., León, K., e Ibáñez, S.J. (2018). Entrenamiento y confiabilidad entre observadores en el análisis del fútbol para ciegos. *Retos: nuevas tendencias en educación física, deporte y recreación*, (34), 155-161. <https://doi.org/10.47197/retos.v0i34.55651>
 28. Garay, J.O., Hernández-Mendo, A., y Morales, V. (2006). Sistema de codificación y análisis de la calidad del dato en el tenis de dobles. *Revista de Psicología del Deporte*, 15(2), 279-294.
 29. Garzón, B. (2010). *Adaptando el baloncesto al niño de 12-13 años: Análisis observacional del lanzamiento del tiro libre* (Tesis doctoral). Universidad de La Rioja, Logroño.
 30. Gorospe, G., Hernández-Mendo, A., Anguera, M. T., y Martínez de Santos, R. (2005). Desarrollo y optimización de una herramienta observacional en el tenis de individuales. *Psicothema*, 17(1), 123-127.
 31. Hernández-Mendo, A., Díaz, F., y Morales, V. (2010). Construcción de una herramienta observacional para evaluar las conductas prosociales en las clases de educación física. *Revista de Psicología del Deporte*, 19(2) 305-318.
 32. Hernández-Mendo, A., y Planchuelo, L. (2012). Una herramienta observacional para la evaluación del desarrollo moral en las clases de educación física en primaria. *Revista Iberoamericana de Psicología del Ejercicio y el Deporte*, 7(2) 287-306.
 33. Jurado, M.C., Anguera, M.T., Reigal, R., y Hernández-Mendo, A. (2021). Análisis de la calidad del dato de un instrumento de observación del clima motivacional: conductas verbales y proxémicas de monitores de mantenimiento físico. *Cuadernos de Psicología del Deporte*, 21(1), 18-42.
 34. Landis, J.R. y Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159-174. <https://doi.org/10.2307/2529310>
 35. Lapresa, D., Álvarez, I., Anguera, M.T., Arana, J., y Garzón, B. (2015). Comparative analysis of the use of space in 7-a-side and 8-a-side soccer: how to determine the minimum sample size in observational methodology. *Journal Motricidade*, 11(4), 92-103. <http://dx.doi.org/10.6063/motricidade.4138>
 36. Lapresa, D., Arana, J., Anguera, M.T., Pérez, J.I., y Amatria, M. (2016). Application of logistic regression models in observational methodology: game formats in grassroots football in initiation into football. *Anales de Psicología*, 32(1), 288-294. <https://doi.org/10.6018/analesps.32.1.186951>
 37. Lapresa, D., Arana, J., Amatria, M., Fernández F.J., y Anguera, M.T. (2017). Fútbol: efectos de una unidad didáctica en la iniciación temprana. *Revista Apunts. Educación Física y Deportes*, 127, 59-67.
 38. Lapresa, D., Arana, J., y Garzón, B. (2020). *Iniciación temprana al fútbol: pautas para orientar la formación*. Barcelona: Fútbol de Libro.
 39. Losada, J.L. y Manolov, R. (2015). The process of basic training, applied training, maintaining the performance of an observer. *Quality & Quantity*, 49(1), 339-347. <https://doi.org/10.1007/s11135-014-9989-7>

40. Mitchell, S.K. (1979). Interobserver agreement, reliability, and generalizability of data collected in observational studies. *Psychological Bulletin*, 86(2), 376-390. <https://doi.org/10.1037/0033-2909.86.2.376>
41. Montesano, A., Feixas, G., y Varlotta, N. (2009). Análisis de contenido de constructos personales en la depresión. *Salud mental*, 32(5), 371-379.
42. Morillo, J.P., Reigal, R.E., y Hernández-Mendo, A. (2015) Análisis del ataque posicional de balonmano playa masculino y femenino mediante coordenadas polares. *RICYDE. Revista Internacional de Ciencias del Deporte*, 11(41), 226-244. <http://dx.doi.org/10.5232/ricyde2015.04103>
43. O'Donoghue, P. (2007). Reliability issues in performance analysis. *International Journal of Performance Analysis in Sport*, 7(1), 35-48. <https://doi.org/10.1080/24748668.2007.11868386>
44. Planes, X. y Anguera, M.T. (2015). Relevancia de los diferentes momentos del juego y de las acciones a balón parado (ABP) en los rendimientos del TC Barcelona y Real Madrid durante la liga nacional de fútbol profesional 2011/12. *Revista Apunts. Educación Física y Deportes*, 121(3), 56-63.
45. Sagastui, J., Herrán, E., y Anguera, M.T. (2020). A Systematic Observation of Early Childhood Educators Accompanying Young Children's Free Play at Emmi Pikler Nursery School: Instrumental Behaviors and Their Relational Value. *Frontiers in Psychology*, 11:1731. <https://doi.org/10.3389/fpsyg.2020.01731>
46. Sastre, V., Lapresa, D., Arana, J., Ibáñez, R., y Anguera, M.T. (2020). Observational analysis of technical-tactical performance in initiation to combat in karate, *International Journal of Performance Analysis in Sport*, 21(1), 126-138. <https://doi.org/10.1080/24748668.2020.1853450>
47. Usabiaga, O., Castellano, J., Blanco-Villaseñor, A., y Casamichana, D. (2013). La Teoría de la Generalizabilidad en las primeras fases del método observacional aplicado en el ámbito de la iniciación deportiva: calidad del dato y estimación de la muestra. *Revista de Psicología del Deporte*, 22(1), 103-109.
48. Usabiaga, O., y Castellano, J. (2014). Uso estratégico del espacio en categorías de formación de pelota vasca. *RICYDE. Revista Internacional de Ciencias del Deporte*, 10(36). <http://dx.doi.org/10.5232/ricyde2014.03602>
49. Vázquez-Diz, J.A., Morillo-Baro, J.P., Reigal, R.E., Morales-Sánchez, V., y Hernández-Mendo, A. (2019a). Contextual factors and decision-making in the behavior of finalization in the positional attack in beach handball: differences by gender through polar coordinates analysis. *Frontiers in Psychology*, 10, 1386. <https://doi.org/10.3389/fpsyg.2019.01386>
50. Vázquez-Diz, J.A., Morillo-Baro, J.P., Reigal, R.E., Morales-Sánchez, V., y Hernández-Mendo, A. (2019b) Mixed methods in decision-making through polar coordinate technique: Differences by gender on beach handball specialist. *Frontiers in Psychology* 10, 1627. doi: 10.3389/fpsyg.2019.01627
51. Vigil-Colet, A., Morales-Vives, F., y Lorenzo-Seva, U. (2013). How social desirability and acquiescence affect the age-personality relationship. *Psicothema*, 25(3), 342-349. <https://doi.org/10.1016/j.paid.2013.07.370>