

Aplicación de la TG en el deporte para el estudio de la fiabilidad, validez y estimación de la muestra¹

Ángel Blanco-Villaseñor*, Julen Castellano**, Antonio Hernández-Mendo***, Carmen Rosa Sánchez-López**** y Oidui Usabiaga**

APPLICATION OF THE GENERALIZABILITY THEORY IN SPORT TO STUDY THE VALIDITY, RELIABILITY AND ESTIMATION OF SAMPLES

KEY WORDS: Generalizability theory, Validity, Reliability, Observational methodology.

ABSTRACT: The generalizability theory (G theory) is a multifaceted error theory that assumes that any measurement situation possesses infinite sources of variation. In the field of observation, it can therefore be applied to determine the influence of these sources of error on measurements. By way of an example, this paper describes three applications of the G theory in the initial stages of observational research: 1) to study validity; 2) to estimate the required sample size; and 3) to study reliability. Currently available software applications (such as GT, EduG or, recently, SAGT) make it easier to conduct this kind of analysis and they provide researchers with procedures for ensuring that the decisions they need to make during the research process are justified prior to being implemented; for instance, calculating the required sample size so as to ensure the accurate generalization of results, or constructing a valid reliable observational instrument that can meet the arduous task of coding and recording behaviour in natural contexts.

En la Psicología del Deporte (PD), en general, es necesario realizar un control que permita una adecuada calidad de los datos procedentes de la Metodología Observacional (MO), así como una mejor y precisa estimación de los diferentes tamaños muestrales. La estructura multivariada de la Teoría de la Generalizabilidad (TG) es la solución a ello (Cronbach, Gleser, Nanda, y Rajaratnam, 1972; Cronbach, Rajaratnam, y Gleser, 1963) y la propuesta de este trabajo. Así, se concretan algunas de las aplicaciones que la TG puede aportar a la MO en el cálculo de la fiabilidad, validez y estimación muestral.

La TG es considerada, por un lado, como una extensión de la Teoría Clásica de los Tests, utilizando los procedimientos del análisis de la varianza y de los diseños experimentales (Martínez, 1995); y, por otro, como una teoría de los errores multifaceta asumiendo que cualquier situación de medida posee infinitas fuentes de variación o facetas (Cronbach, Gleser, Nanda y Rajaratnam, 1972). A través de la TG se pueden analizar las diferentes fuentes de variación que pueden estar afectando a una medida o diseño de medida de origen observacional. La aplicación de esta teoría permite estimar el grado de generalización de un diseño de medida con respecto a las condiciones particulares de un valor teórico buscado. El coeficiente de generalizabilidad permite estimar el ajuste de la media observada a la media de todas las observaciones posibles.

La TG unifica las definiciones de fiabilidad, validez y precisión. Estas corresponden a un aspecto parcial de un modelo

más general, que considera todas aquellas fuentes de variación que afectan a los resultados observados. Uno de los objetivos de la medición es identificar y medir los componentes de variancia que aportan error a una estimación e implementar estrategias que reduzcan la influencia de estas fuentes de error sobre la medida (Blanco-Villaseñor, 1993; Blanco-Villaseñor, Castellano y Hernández-Mendo, 2000). El cumplimiento de este objetivo pasa ineludiblemente por la estimación de los componentes de variancia de diversas facetas o fuentes de variabilidad que componen el estudio observacional. El investigador definirá el conjunto de condiciones que restringen las facetas. El conjunto de valores que puede tomar una faceta se conoce como universo de generalización. El estudio de Generalizabilidad se compone básicamente de cuatro fases: (1) Definición de las facetas de estudio; (2) Análisis de variancia de las puntuaciones obtenidas sobre las facetas de estudio; (3) Cálculo de los componentes de error; (4) Optimización de los coeficientes de Generalizabilidad.

En las ciencias del comportamiento, los fenómenos observados están influidos por tal cantidad de factores que la repetición de una misma experiencia o la utilización de cualquier otro instrumento puede modificar considerablemente el resultado que se obtuvo la primera vez. Por ello, la actitud científica más elemental lleva a preguntarse si esos valores observados son interpretables o si, por el contrario, son el resultado de fluctuaciones aleatorias, introducidas por la propia medida (Blanco-Villaseñor, 1989, 1991, 1992, 1993). Este interrogante

Correspondencia: Antonio Hernández-Mendo. Dpto. Psicología Social, A.T., T.S. y S.S. Facultad de Psicología, Universidad de Málaga. Campus de Teatinos, s/n. 29071 Málaga (Spain). E-mail: mendo@uma.es

¹ Este trabajo forma parte de la investigación Observación de la interacción en deporte y actividad física: Avances técnicos y metodológicos en registros automatizados cualitativos-cuantitativos, que ha sido subvencionado por la Secretaría de Estado de Investigación, Desarrollo e Innovación del Ministerio de Economía y Competitividad [DEP2012-32124], durante el trienio 2012-2015. Este trabajo se inscribe en el Grupo de Investigación Consolidado de Cataluña GRUPO DE INVESTIGACIÓN E INNOVACIÓN EN DISEÑOS (GRID). Tecnología y aplicación multimedia y digital a los diseños observacionales, que ha sido subvencionado por el Departamento de Universidades, Investigación y Sociedad de la Información de la Generalitat de Catalunya [2009 SGR 829] durante el periodo 2009-2013.

* Universidad de Barcelona.

** Universidad del País Vasco.

*** Universidad de Málaga.

**** Universidad de La Laguna.

— Artículo invitado con revisión.

es particularmente necesario en los diseños de observación de la conducta (Blanco-Villaseñor y Anguera, 2003).

Con todo, el objetivo del presente trabajo es presentar tres aplicaciones de la TG en sendas fases de una investigación en el ámbito observacional: 1) estudio de la validez, 2) estimación de la muestra; y, 3) estudio de la fiabilidad.

1. Aplicación práctica para el estudio de la validez

Los diferentes enfoques para evaluar y describir la validez pueden reducirse a la validez convergente, de criterio o concurrente (cuando se dispone de mediciones de idénticos o similares atributos) y validez de constructo (cuando no se dispone de otra medida) (Blanco-Villaseñor y Anguera, 2003). El concepto de validez sufrió, a partir de los años 90, una importante transformación conceptual gracias al trabajo de Messick (1989) al definir validez como el grado de propiedad de las inferencias e interpretaciones que se derivan de las puntuaciones de los instrumentos de medida, incluyendo las consecuencias sociales que resultan de su aplicación (Padilla, Gómez, Hidalgo y Muñiz, 2006), lo cual es aplicable a los instrumentos estándar. Así, en vez de citar distintos tipos de validez, Messick indica la recogida de diferentes tipos de evidencias, de acuerdo con los propósitos y usos de dichos instrumentos, entre ellas, evidencias de contenido, predictivas y de constructo. Messick considera la validez, no un rasgo dicotómico, sino una cuestión de grado, no se puede afirmar que una prueba es válida, sino más propiamente que la prueba exhibe un grado aceptable de validez para ciertos usos específicos y con ciertas poblaciones (Messick, 1989; Padilla et al., 2006).

La validez psicométrica de un instrumento es sólo una parte de la sistemática y rigurosa recolección de evidencia empírica que debe emprenderse cuando se realiza la pregunta: ¿Son apropiadas las inferencias generadas a partir de las puntuaciones obtenidas a través de una prueba? Este último concepto de validez será de suma importancia en la observación.

La literatura científica presenta diversos ejemplos (inseparables del cálculo de la fiabilidad) en fútbol (Blanco-Villaseñor, Castellano y Hernández-Mendo, 2000), fútbol-7 (Reina, Hernández-Mendo y Fernández-García, 2009), tenis (Gorospe, Hernández-Mendo, Anguera y Martínez de Santos, 2005), tenis de dobles (Garay Plaza, Hernández-Mendo y Morales Sánchez, 2006), voleibol (Hernández-Mendo, Montoro, Reina y Fernández-García, 2012), pelota vasca (Usabiaga, Castellano, Blanco-Villaseñor y Casamichana, 2013), desarrollo moral (Hernández-Mendo, Díaz Martínez y Morales Sánchez, 2010; Hernández-Mendo y Planchuelo, 2012) e intervención inclusiva en educación física (Fernández, Sánchez, Jiménez, Navarro y Anguera, 2012). En este último estudio encontramos un claro ejemplo de utilización de la TG en el ámbito de la validez.

Procedimiento

Para dicho estudio se elaboró un sistema de categorías específico a partir de tres situaciones que marcan el grado de inclusión del alumnado con discapacidad motriz (DM) en educación física (EF), esto es: exclusión, integración e inclusión (que conformarían los criterios), seis categorías y 44 subcategorías (Fernández et al., 2012)

Las unidades de observación empleadas son muestreadas a través de un registro continuo (Anguera, 1993, p.15). La lógica didáctica con la que se proponen y desarrollan las tareas motrices y su alcance inclusivo ha constituido el referente principal para

el diseño de las categorías. De igual modo, atendiendo al proceso en un contexto de interacción (Castellano y Hernández-Mendo, 2003) destacamos su valor sincrónico y diacrónico, pues permiten seguir el proceso de intervención (Anguera, Blanco-Villaseñor, Hernández-Mendo y Losada, 2011). El estudio se realizó en tres centros preferentes de integración de alumnado con DM de Santa Cruz de Tenerife. Han participado seis profesores de EF y 171 alumnos de ESO, incluidos 10 alumnos con DM.

Se filmaron 10 sesiones de clases de EF con una duración media de 40 minutos cada una, cinco correspondientes a la unidad didáctica de la fase inicial o diagnóstico (fase A) y las otras cinco correspondientes a la fase final (fase B). Entre dichas fases se produce el proceso colaborativo de investigación-acción donde se promueve la utilización de estrategias inclusivas en el desarrollo de las clases de EF.

Resultados

En primer lugar se valoró la fiabilidad inter-observadores utilizándose para ello el índice Kappa de Cohen ($\kappa = .80$). Para el control de la calidad del dato (Blanco-Villaseñor y Hernández-Mendo, 1998; Blanco-Villaseñor y Anguera, 2003; Sánchez, Jiménez y Navarro, 2008; Sánchez, Blanco-Villaseñor y González, 2010) se aplicó la TG (Cronbach, Gleser, Nanda y Rajaratnam, 1972) que permite evaluar las distintas fuentes de variabilidad (facetas), pudiéndose determinar qué dimensión debe tener cada una de ellas para obtener puntajes de alta precisión.

En cuanto a la valoración de la generalizabilidad obtenida por las diez sesiones observadas, las seis categorías y las 44 subcategorías que tiene el sistema, se planteó un diseño de tres facetas: sesión (S) categoría (C) subcategoría (X), estudiándose los tres modelos posibles: CX/S, XS/C y SC/X. El Plan de Optimización determinó que sería necesario observar 28 sesiones para obtener una generalizabilidad de 0.91; 8 categorías para obtener una generalizabilidad de 0.90; pero que, con las 44 subcategorías que contiene el sistema, la generalizabilidad obtenida es de 0.99.

Con respecto a la validez, se estructuró un diseño de cinco facetas: ciclo (Y), centro (C), caso (D), fase (F) y criterio (R), que permitiera estimar la validez de los criterios utilizados en el sistema de categorías. Los coeficientes relativos y absolutos de la faceta criterio, en este caso fueron 0.00. La validez debe obtener un coeficiente de generalizabilidad próximo a 0, indicativo de la heterogeneidad del sistema y del requisito de exhaustividad y mutua exclusividad (E/ME), por lo tanto el valor estimado es óptimo.

Discusión

Los resultados indican que el sistema de categorías empleado en la muestra es fiable y válido para evaluar la intervención educativa que posibilite la inclusión de alumnado con DM.

2. Aplicación práctica para el estudio de la fiabilidad

La fiabilidad evalúa que una medición determinada pueda ser reproducible; no da información acerca de la naturaleza de la medida (Blanco-Villaseñor y Anguera, 2003). El análisis de la fiabilidad es una de las cuestiones de capital importancia dentro de la MO al abordar el análisis de calidad del dato. En este punto se hace vital estimar en qué medida los observadores son fiables, mediante el análisis de la estabilidad de la medida intrasesión (modelos C/O - categorías/observadores) o de la estabilidad intersesión (modelos S/O, sesiones/observadores ó SC/O, sesiones*categorías/observadores).

Existen diversas evidencias de la utilización de los análisis de generalizabilidad para la estimación de la fiabilidad utilizando la MO en el deporte (Blanco-Villaseñor, Castellano y Hernández-Mendo, 2000; Castellano, Hernández Mendo, Gómez de Segura, Fontetxa y Bueno, 2000; Garay Plaza, Hernández Mendo y Morales Sánchez, 2006; Gorospe, Hernández-Mendo, Anguera y Martínez de Santos, 2005; Hernández-Mendo, Montoro Escaño, Reina Gómez y Fernández-García, 2012; Hernández-Mendo y Planchuelo, 2012; Reina-Gómez, Hernández-Mendo y Fernández García, 2009;

Hernández-Mendo, Díaz Martínez y Morales Sánchez, 2010;) e incluso en la clínica deportiva (Cuesta-Vargas y Hernández-Mendo, 2009).

Procedimiento

Un ejemplo de la utilización de la TG para el estudio de la fiabilidad es el presentado en el trabajo Hernández-Mendo, Díaz Martínez y Morales Sánchez (2010). En este trabajo se construye un sistema mixto constituido por cinco criterios y 37 categorías (tabla 1).

Criterio 1. Conductas hacia sus Iguales	
<i>1.1. Macro categoría conductas positivas hacia sus iguales</i>	
A	Asiste, auxilia a los demás participantes.
I	Introduce a los participantes en la tarea.
CO	Colabora en el desarrollo de la tarea con los demás participantes.
S	Busca una mejora de nivel o perfeccionamiento con la práctica repetida.
AP	Anima a los demás participantes durante el desarrollo de la tarea.
DM	Participa con claro autocontrol en la tarea.
R	Desarrolla de forma adecuada su implicación en el aprendizaje.
<i>1.2. Macro categoría conductas negativas hacia sus iguales</i>	
NDM	Falta de autocontrol en la tarea.
AV	Ofende con gestos o acciones a los demás participantes.
NI	No introduce a los participantes en la tarea.
NCO	No colabora en el desarrollo de la tarea, con los demás participantes.
IN	Molesta a los demás participantes durante el desarrollo de las tareas.
NR	No desarrolla de forma adecuada su implicación en el aprendizaje.
Criterio 2. Conductas hacia el docente	
<i>2.1. Macro categoría conductas positivas hacia el docente</i>	
PA	Presta atención en las explicaciones del docente.
AS	Colabora con el docente en la organización de la tarea.
PAR	Participa de forma activa en la tarea.
<i>2.2. Macro categoría conductas negativas hacia el docente</i>	
NPA	No presta atención en las explicaciones del docente.
NAS	No colabora con el docente en la organización de la tarea.
DIS	Lleva a cabo conductas disruptivas con el docente.
EFP	Lleva a cabo conductas de provocación con el docente.
NPAR	No desarrolla las tareas de forma inmediata, se inhibe en las mismas.
Criterio 3. Conductas hacia equipamiento y material	
<i>3.1. Macro categoría conductas positivas hacia el material/instalaciones</i>	
CI	Conserva y conoce las instalaciones donde se desarrolla la tarea.
CVRG	Utiliza el material de forma debida, lo ubica de forma adecuada.
<i>3.2. Macro categoría conductas negativas hacia el material/instalaciones</i>	
NCI	No conserva ni conoce las instalaciones donde se desarrolla la tarea.
NCVRG	No utiliza el material de forma debida, no lo ubica adecuadamente.
Criterio 4. Conductas hacia reglas-normativa	
<i>4.1. Macro categoría conductas positivas hacia las reglas/normativa</i>	
RCR	Acepta y reconoce el reglamento del deporte/actividad a desarrollar.
PT	Para la tarea para auxiliar, asistir a un participante lesionado.
RT	Retoma la tarea en la situación en la que se detuvo.
IT	Identifica y lleva a cabo acciones de estrategia, fundamentales en la tarea.
<i>4.2. Macro categoría conductas negativas hacia las reglas/normativa</i>	
NRCR	No acepta el reglamento del deporte/actividad a desarrollar.
NPT	No para la tarea para auxiliar, asistir a un participante lesionado.
NRIT	No retoma la tarea en la situación en la que se detuvo, para obtener ventaja.
JP	Realiza acciones peligrosas con su conducta.
PTI	Deja pasar el tiempo deliberadamente.
NIT	No lleva a cabo acciones de estrategia, fundamentales en la tarea.
Criterio 5. Faltas por contacto	
FG	Conducta antirreglamentaria por juego duro.
FL	Conducta antirreglamentaria por juego brusco.

Tabla 1. Criterios, Macro categorías y Categorías de la Herramienta de Observación.

La herramienta de observación se construye a partir de la interacción de las conductas con las tareas, que puede ser de tres tipos, individual, de cooperación, y de cooperación-oposición. Esta interacción se puso en correspondencia con los diferentes niveles de los estadios de desarrollo moral propuestos por Kohlberg (1992).

Las unidades de observación se corresponden a multieventos (Bakeman y Quera, 1996). El registro de los datos es continuo. El procedimiento de registro fue la *codificación directa* a través de la hoja de cálculo Excel XP, realizada por dos equipos de observación compuestos cada uno de ellos por tres observadores, entrenados previamente a partir de un protocolo observacional. Se emplearon tres sesiones de 30 minutos de duración cada una, correspondientes a la parte central de la sesión de EF, no realizándose observaciones en el calentamiento y en la vuelta a la calma.

Cada uno de los grupos utilizó la concordancia consensuada (Anguera, 1990), en todas las ocasiones. Uno de los dos grupos de observadores realizó dos veces la misma observación, con un periodo intermedio de 15 días, con el objetivo de realizar posteriormente un análisis de la fiabilidad intraobservadores.

La inobservabilidad estimada es de tipo tecnológico (Hernández-Mendo, 1996), y en ningún caso ha superado la ruptura de la continuidad de la sesión de observación por un

periodo superior al 10% del total de la misma (Anguera, 1990), por lo cual no se ha estimado su influencia en la presente investigación. En cuanto a los requisitos de constancia *intra-intersesional*, se garantizó al máximo, escogiéndose sesiones donde no existían circunstancias excepcionales de ruptura del flujo conductual. Y en cuanto a la constancia intersesional, además de ser las sesiones propias del contenido didáctico de deportes colectivos (el Voleibol) en su desarrollo se han seguido los objetivos, contenidos, metodologías propias del contexto educativo.

Resultados

En el análisis de calidad del dato se realizó un estudio de correlaciones que resulta satisfactorio situándose entre el .98 y 1.00; tanto para la *concordancia intra-observadores* (Pearson = 1.00, Tau Kendall = .98, Spearman = .99) como para la *concordancia inter-observadores* (Pearson = 1.00, Tau Kendall = .98, Spearman = .99).

Se realizó una estimación del índice *kappa* de Cohen (1960). Este índice es un estadístico de concordancia que corrige el azar. Se ha utilizado para estimar la fiabilidad por criterios y el global de la sesión, siendo de .91 para la *concordancia intra-observadores*, y de 1.00 para la *concordancia inter-observadores*.

	Suma Cuadrados	gl	Media Cuadrados	Error Estandar	%
Observadores [O]	0.55	2	273	0.00557	0
Categorías [C]	281231.10	36	7811975	697.39666	100
CO	31.47	72	437	0.07185	0
Índice de Fiabilidad	1.00				
Índice de Generalizabilidad	1.00				

Tabla 2. Fiabilidad inter-observadores para el diseño C/O.

En el cálculo de la *fiabilidad inter-observadores* se empleó un diseño de dos facetas (categorías/observadores = C/O) (ver tabla 2). La determinación de las fuentes de varianza reveló que la mayor parte de la variabilidad (100%), quedaba asociada a la faceta categorías, siendo nula en el resto de facetas, *observadores*, y *observadores/categorías* (0%). El análisis de los coeficientes de generalizabilidad obtuvo resultados que indicaban una alta fiabilidad en la precisión de generalización de los resultados (1.00).

Discusión

La importancia de este trabajo, se centra en la construcción y validación de una herramienta *ad hoc* utilizando fundamentalmente análisis de generalizabilidad. Los coeficientes de generalizabilidad han sido satisfactorios, especialmente los índices de fiabilidad y generalizabilidad. Se verifica que el sistema mixto propuesto cumple con los requisitos metodológicos relativos a la fiabilidad, precisión y validez exigibles. Se puede, así, obtener información de los diferentes aspectos de las conductas de las clases de EF, conductas relacionadas con el desarrollo positivo.

3. Aplicación práctica para la estimación de la muestra

Una vez resuelta la decisión entorno al muestreo observacional (Anguera, Blanco-Villaseñor, Hernández Mendo y Losada, 2011), uno de los siguientes pasos es determinar el número de sesiones de observación (partidos, sets, cuartos, *rounds*, etc...) necesarias para generalizar con precisión. Éste queda en manos del investigador, pues deberá evaluar el coste/beneficio de una muestra mayor que pueda mejorar substancialmente la precisión en la generalización de los datos. Este aspecto ha sido abordado en diversos trabajos (Casamichana, Castellano, Blanco-Villaseñor, y Usabiaga, 2012; García García, Hernández-Mendo, Serrano y Morales Sánchez, 2013; Usabiaga, Castellano, Blanco-Villaseñor y Casamichana, 2013).

En el estudio propuesto por Usabiaga, Castellano, Blanco-Villaseñor y Casamichana (2013), contextualizado en el deporte de formación, uno de los problemas de investigación que se pretendió resolver fue estimar la precisión de generalización de los partidos codificados de pelota vasca con una herramienta de observación.

En dicho trabajo se registraron cuatro partidos de pelota, de los cuales dos fueron de la categoría benjamín (8-10 años) y los otros dos de la categoría alevín (10-12 años), siendo dos partidos de la temporada 2007-2008 y otros dos de la temporada 2008-2009.

Para la codificación de la acción de juego se utilizó la herramienta de observación EBSIS.e (Usabiaga y Castellano, 2005, 2011) configurada por una combinación de formatos de campo y sistemas de categorías. La herramienta taxonómica está constituida por cinco criterios y 100 categorías, que describen la actividad de los deportistas considerando la propia interacción de la acción de juego. Una vez grabados los partidos, éstos fueron visionados, codificados y registrados usándose el software para la observación deportiva MOTS (Castellano, Perea y Hernández-Mendo, 2008).

Procedimiento

Una vez que los partidos están codificados y registrados, se llevó a cabo un estudio D (Decisión). Para ello se diseñó un modelo de cuatro facetas (niveles, partidos, criterios y categorías) con el siguiente plan de medida, NP/(K:C), donde N son los niveles (2), P son los partidos (2), K son las conductas o categorías (100) y C son criterios (5), estando las categorías anidadas en la faceta criterio (K:C). Para llevar a cabo los análisis estadísticos se utilizaron los programas *EduG* versión 6.0 (Cardinet, Johnson y Pini, 2010) y el *SAS v.9.1.* (Schlotzhauer y Littell, 1997; SAS InstituteInc., 1999).

Resultados

Previo al estudio D se realizó un estudio de generalizabilidad (G) donde se estimaron los porcentajes de varianza de cada una de las facetas y sus interacciones, así como los coeficientes absolutos y relativos de generalizabilidad recogidos en la tabla 3.

Facetas	gl	Pr > F para el modelo <.0001		
		SC (Tipo III)	Pr > F	% de varianza
N	1	3287.29	.0002	0
P	1	8308.08	<.0001	0
N*P	1	8541.83	<.0001	6.5
K:C	99	331124.55	<.0001	63.2
N*K:C	99	22345.80	.50	0
P*K:C	99	33288.34	.004	0
N*P*K:C	99	37419.10	.0003	30.2

N K:C / P $\xi\rho^2(\delta) = .77$ y $\xi\rho^2(\Delta) = .77$

Tabla 3. Valores del coeficiente de determinación (r^2), gl, suma de cuadrados (SC tipo III) y porcentaje de varianza de facetas e interacciones y coeficientes absolutos y relativos de generalizabilidad ($\xi\rho^2(\delta)$ y $\xi\rho^2(\Delta)$) para el diseño de cuatro facetas N*P*K:C.

En relación a los análisis del estudio G, la precisión de generalización para estos cuatro partidos codificados (dos categorías y dos años), fue moderada (0.77). Llegados a este punto se realizó el estudio D para estimar de manera *apriorística* el número de partidos necesarios para estimar con precisión. En la Tabla 4 se recogen los valores de precisión, coeficientes de

generalización relativos y absolutos, estimados a partir de la codificación de los dos primeros partidos de cada año y categoría. Como puede comprobarse, los valores se aproximaron cada vez más a la unidad, fueron por tanto más precisos, a medida que se incluyeron un mayor número de partidos a codificar en cada uno de los niveles.

2*	4	6	8	10	15
$\xi\rho^2(\delta) = 0.77$	$\xi\rho^2(\delta) = 0.87$	$\xi\rho^2(\delta) = 0.91$	$\xi\rho^2(\delta) = 0.93$	$\xi\rho^2(\delta) = 0.95$	$\xi\rho^2(\delta) = 0.96$
$\xi\rho^2(\Delta) = 0.77$	$\xi\rho^2(\Delta) = 0.87$	$\xi\rho^2(\Delta) = 0.91$	$\xi\rho^2(\Delta) = 0.93$	$\xi\rho^2(\Delta) = 0.95$	$\xi\rho^2(\Delta) = 0.96$

*Estimación a partir de 2 partidos por nivel

Tabla 4. Valores estimados de los coeficientes relativos ($\xi\rho^2(\delta)$) y absolutos ($\xi\rho^2(\Delta)$) de generalizabilidad para el plan de optimización del modelo [NK:C/P] para una estimación del número de partidos.

Discusión

Se han expuesto los pasos y las interpretaciones que se realizan para estimar apriorísticamente la muestra necesaria para llevar a cabo el proyecto de investigación que el investigador quiere resolver. Con los valores de la tabla 4, se dispone de información que permitirá decidir (de ahí la denominación de estudio D) el número de sesiones de observación que deberá incluir en su trabajo, utilizando un análisis coste/beneficio, para que la precisión de generalización escogida respecto al número de partidos codificados se ajuste a sus necesidades. Las diversas posibilidades que nos ha permitido la TG en Psicología del Deporte a través de la MO tiene aún más opciones, como la aproximación *apriorística versus* aplicaciones *aposteriori*. En este sentido, cualquiera de los trabajos presentados pueden ser

considerados desde un punto de vista *a priori* como una exploración de un dominio de la investigación insuficientemente conocido o estudiado, y suponen una forma de preparar investigaciones a mayor escala. También es posible que los tamaños muestrales puedan ajustarse mejor a la realidad social de las investigaciones en psicología de la actividad física y el deporte.

Se pueden confirmar las expectativas de investigación mediante una aplicación a posteriori de cualquiera de los trabajos aquí presentados. En ese sentido, si las exploraciones previas sirven para tomar decisiones en una investigación a mayor escala, también pueden servir para confirmar los resultados. En cualquier caso, se trata de algunas de las posibilidades que mejorarán la calidad de investigaciones y de publicaciones científicas, que contarán con mayor rigor y precisión.

APLICACIÓN DE LA TG EN EL DEPORTE PARA EL ESTUDIO DE LA VALIDEZ, FIABILIDAD Y ESTIMACIÓN DE LA MUESTRA

PALABRAS CLAVE: Teoría de la Generalizabilidad, Validez, Fiabilidad, Metodología observacional.

RESUMEN: La Teoría de la Generalizabilidad (TG) es una teoría de los errores multifaceta que asume que cualquier situación de medida posee infinitas fuentes de variación. En este sentido esta teoría puede aplicarse en el ámbito de la observación con la intención de conocer la influencia de estas fuentes de error sobre la medida. De todas las posibilidades, en este trabajo se presentan, a modo de ejemplo, tres aplicaciones de la TG en las primeras fases de una investigación en el ámbito observacional: 1) para el estudio de la validez, 2) para la estimación de la muestra; y, 3) para el estudio de la fiabilidad. Actualmente existen aplicaciones informáticas (GT, EduG o, recientemente, SAGT) que facilitan la implementación de este tipo de análisis que permiten al investigador contar con recursos procedimentales para que las decisiones que tiene que tomar en el proceso de investigación puedan estar justificadas antes de ser implementadas como la muestra necesaria para poder generalizar con precisión o disponer de una herramienta de observación válida y fiable para afrontar el arduo proceso de codificación y registro de las conductas acontecidas en los contextos naturales donde se aplican.

APLICAÇÃO DA TG NO DESPORTO PARA O ESTUDO DA VALIDADE, FIDELIDADE E ESTIMAÇÃO DA AMOSTRA

PALAVRAS-CHAVE: Teoria da Generalização, validade, fidelidade, metodologia observacional.

RESUMO: A Teoria da Generalização (TG) é uma teoria de erros multifacetada que assume que qualquer situação de medida possui infinitas fontes de variação. Neste sentido, esta teoria pode ser aplicada no âmbito da observação com a intenção de conhecer a influência destas fontes de erro sobre a medida. De todas estas possibilidades, neste trabalho apresentam-se, a título de exemplo, três aplicações da TG nas primeiras fases de uma investigação no âmbito observacional: 1) para o estudo da validade, 2) para a estimativa da amostra; e, 3) para o estudo da fidelidade. Actualmente existem aplicações informáticas (GT, EduG ou, recentemente, SAGT) que facilitam a implementação deste tipo de análise que permitem ao investigador contar com recursos procedimentais para que as decisões que tem que tomar no processo de investigação possam ser justificadas antes de serem implementadas como a amostra necessária para poder generalizar com precisão ou dispor de uma ferramenta de observação válida e fiável para levar a cabo o processo de codificação e registo dos comportamentos, nos contextos naturais onde decorrem.

Referencias

- Anguera, M. T. (1990). Metodología observacional. En J. Arnau, M. T. Anguera y J. Gómez. *Metodología de la investigación en Ciencias del Comportamiento* (pp. 125-236). Murcia: Secretariado de Publicaciones de la Universidad de Murcia.
- Anguera, M. T. (Ed.) (1993). *Metodología observacional en la investigación psicológica*. 2 (vols.). Barcelona: PPU.
- Anguera, M. T., Blanco-Villaseñor, A., Hernández-Mendo, A. y Losada, J. L. (2011). Diseños observacionales: ajuste y aplicación en psicología del deporte. *Cuadernos de Psicología del Deporte*, 11(2), 63-76.
- Blanco-Villaseñor, A. (1989). Fiabilidad y generalización de la observación conductual. *Anuario de Psicología*, 43(4), 5-32.
- Blanco-Villaseñor, A. (1991). La Teoría de la Generalizabilidad aplicada a diseños observacionales. *Revista Mexicana de Análisis de la Conducta*, 14(3), 23-64.
- Blanco-Villaseñor, A. (1992). Aplicaciones de la Teoría de la Generalizabilidad en la selección de diseños evaluativos. *Bordón*, 43(4), 431-459.
- Blanco-Villaseñor, A. (1993). Fiabilidad, precisión, validez y generalización de los diseños observacionales. En M. T. Anguera (Ed.), *Metodología observacional en la investigación psicológica* (Vol 2: Fundamentación, pp. 151-261). Barcelona: PPU.
- Blanco-Villaseñor, A. y Anguera, M. T. (2003). Calidad de los datos registrados en el ámbito deportivo. En A. Hernández-Mendo, *Psicología del Deporte (Vol. II): Metodología* (pp. 35-73). Buenos Aires: Tulio Guterman (www.efdeportes.com).
- Blanco-Villaseñor, A., Castellano, J. y Hernández-Mendo, A. (2000). Generalizabilidad de las observaciones de la acción del juego en el fútbol. *Psicothema*, 12(2), 81-86.
- Blanco-Villaseñor, A. y Hernández-Mendo, A. (1998). Estimación y generalización en un diseño de estructura espacial. En J. Sabucedo, R. García, E. Ares y D. Prada, *Medio ambiente y responsabilidad humana* (pp. 579-583). A Coruña: Libro de Comunicaciones. VI Congreso de Psicología Ambiental.
- Cardinet, J., Johnson, S. y Pini, G. (2010). *Applying Generalizability Theory using EduG*. Nueva York: Routledge Academic.
- Casamichana, D., Castellano, J., Blanco-Villaseñor, A. y Usabiaga, O. (2012). Estudio de la percepción subjetiva del esfuerzo en tareas de entrenamiento en fútbol a través de la teoría de la TG. *Revista de Psicología del Deporte*, 21(1), 35-40.
- 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), 636-641.
- 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.
- Cronbach, L. J., Rajaratnam, N. y Gleser, G. C. (1963). Theory of generalizability: a liberalization of reliability theory. *British Journal of Mathematical and Statistical Psychology*, 16, 137-163.
- 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.
- Cuesta-Vargas, A. I. y Hernández-Mendo, A. (2009). Fiabilidad de un sistema de clasificación clínica de lumbalgia mecánica inespecífica para el tratamiento de la estabilidad lumbo pélvica mediante el control motor. *Cuest. fisioter*, 38(1):11-18.
- Fernández, J. M., Sánchez, C. R., Jiménez, F., Navarro, V. y Anguera, M. T. (2012). Sistema de codificación y análisis de la calidad del dato para una intervención inclusiva en educación física. *Revista de Psicología del Deporte*, 21(1), 103-109.
- Garay Plaza, J. O., Hernández-Mendo, A. y Morales Sánchez, 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.
- García García, O., Hernández-Mendo, A., Serrano, V. y Morales Sánchez, V. (2013). Aplicación de la teoría generalizabilidad a un análisis de tensiomiografía en ciclistas profesionales de ruta. *Revista de Psicología del Deporte*, 22(1), 53-60.
- 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.
- Hernández-Mendo, A., Díaz, F. y Morales Sánchez (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.
- Hernández-Mendo, A., Montoro, J., Reina, A. y Fernández-García, J.C. (2012). Desarrollo y optimización de una herramienta observacional para el bloqueo en voleibol. *Revista Iberoamericana de Psicología del Ejercicio y el Deporte*, 7(1), 15-32.
- 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.
- Kohlberg, L. (1992). *Psicología del desarrollo moral*. Bilbao: Desclée de Brouwer.
- Martínez Arias, R. (1995). *Psicometría: teoría de los test psicológicos y educativos*. Madrid: Síntesis.
- Messick, S. (1989). Validity. En R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). Nueva York: Macmillan.
- Padilla J. P., Gómez, J., Hidalgo, M. D. y Muñoz, J. (2006). La evaluación de las consecuencias del uso de los tests en la teoría de validez. *Psicothema*, 18(2), 307-312.
- Reina, A., Hernández-Mendo, A. y Fernández-García, J. C. (2009). Multi-facet design for goal scoring in SOCCER-7. *Quality & Quantity*, 44(5), 1025-1035.
- Sánchez, C. R. Fernández, J. M. y Jiménez, F. (2008). Teoría de la Generalizabilidad en intervención docente de Educación Física. *II Congreso de la Sociedad Iberoamericana de Psicología del Deporte. 6-8 noviembre*. Torrelavega, España.
- Sánchez, C. R., Blanco-Villaseñor, A. y González, J. (2010). Optimización de un modelo para baloncesto mediante Teoría de la Generalizabilidad. *XII Congreso Nacional de Psicología de la Actividad Física y el Deporte. 23-26 junio*. Madrid, España.
- 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.
- Usabiaga, O. (2005). *Evaluación de la acción de juego de la pelota vasca: aplicación en mano parejas*. Tesis doctoral no publicada. Universidad del País Vasco UPV/EHU.
- Usabiaga, O. y Castellano, J. (2011). Adaptación de la herramienta de observación de la pelota a mano EBSIS para el ámbito formativo. *Lecturas: Educación Física y Deportes*, 16(162). <http://www.efdeportes.com/efd162/observacion-de-la-pelota-a-mano-ebis.htm>. [Consulta: 23 junio 2013].