

SAGT: APLICACIÓN INFORMÁTICA PARA ANÁLISIS DE GENERALIZABILIDAD

Antonio Hernández-Mendo¹, Ángel Blanco-Villaseñor², José Luis Pastrana¹, Verónica Morales-Sánchez¹, Francisco Jesús Ramos-Pérez¹

Universidad de Málaga¹, España y Universidad de Barcelona², España

RESUMEN: Se presenta un nuevo software para realizar análisis de generalizabilidad. Las características del mismo superan las limitaciones de programas anteriores (GENOVA, TG o EduG), tanto a nivel de tratamiento numérico como gráfico y en la importación/exportación de datos. El entorno del programa es más amigable. Para demostrar su eficacia se utiliza en el análisis de datos y valoración de resultados de un programa de intervención para la reducción de discriminación de género en las clases de educación física (EF). Se utiliza un diseño observacional y un diseño de generalizabilidad ortogonal con 6 facetas parcialmente anidado. El programa se desarrolla durante seis meses. La muestra está compuesta por 100 alumnos (51 chicas y 49 chicos) con edades comprendidas entre los 10 y los 12 años que cursaban 5º de Educación Primaria y se encontraban distribuidos en cuatro grupos. Los resultados del análisis de generalizabilidad señalan que la herramienta de observación es fiable, válida y precisa y que los observadores son altamente fiables ($k = 0.73-0.81$, $G = 1$). El ANOVA realizado señala que el programa de intervención ha sido eficaz. Además, se ha demostrado la eficacia y usabilidad del programa SAGT.

PALABRAS CLAVE: Generalizabilidad, observación.

SAGT: NEW SOFTWARE FOR GENERALIZABILITY ANALYSIS

ABSTRACT: This paper describes a new software application for generalizability analysis. The characteristics of this software overcome the limitations of existing programs (GENOVA, GT and EduG) at both the numerical and graphical level, as well as in relation to data import/export. Its environment is also more user-friendly. The effectiveness of the new software is illustrated by analysing data and evaluating results from an intervention programme whose aim was to reduce gender discrimination in physical education classes. This part of the study used an observational design and a partially-nested orthogonal generalizability design with six facets. The intervention programme lasted six months and involved 100 children aged 10-12 years (51 girls and 49 boys) who were distributed across four classes in primary year five. The results of the generalizability analysis indicated that the observation instrument was a reliable, valid and precise tool, and also that observers were highly reliable ($k = 0.73-0.81$, $G = 1$). The ANOVA carried out indicated that the intervention programme had been effective. The SAGT software was likewise shown to be effective and user-friendly.

KEYWORDS: reasons, physical activity, physical self-concept, extracurricular activities.

SAGT: SOFTWARE PARA ANÁLISE DE GENERALIZAÇÃO

RESUMO: Ele apresenta um novo software para análise de generalização. Suas características superar as limitações dos programas anteriores (Génova, ou EDUG TG), tanto em termos de processamento numérico e gráfico e dados de importação/exportação. O ambiente do programa é mais amigável. Para demonstrar a sua eficácia utilizados na análise e avaliação dos resultados de um programa de intervenção para reduzir a discriminação de género em aulas de educação física (EF). Um projeto observacional e generalização projeto ortogonal 6 facetas aninhados parcialmente utilizado. O programa é executado por seis meses. A amostra é constituída por 100 alunos (51 meninas e 49 meninos) com idade entre 10 e 12 anos que estavam no Ensino Primário e 5 foram divididos em quatro grupos. Os resultados da análise indicam que a generalização da ferramenta de observação é fiável, preciso e válidos e que os observadores são altamente fiáveis ($k = 0.73$ para 0.81 , $G = 1$). A ANOVA realizada indica que o programa de intervenção tem sido eficaz. Também demonstrou a eficácia ea usabilidade do programa sagt.

PALAVRAS-CHAVE: Generalização, observação.

Manuscrito recibido: 08/09/2015
Manuscrito aceptado: 20/09/2015

Dirección de contacto: Antonio Hernández Mendo Dpto. Psicología Social, Trabajo Social, Antropología Social y Estudios de Asia Oriental. Campus Teatinos s/n. 29071 Málaga, España.
Correo-e: mendo@uma.es

La *Teoría de la Generalizabilidad* (TG) es relativamente joven en cuanto a su uso. Puede considerarse una extensión de la Teoría Clásica de los Tests (TCT), utilizando los procedimientos del Análisis de la Varianza y de los diseños experimentales (Martínez-Arias, 1995). Además, es una teoría de los errores multifaceta, asume que cualquier situación de medida posee infinitas fuentes de variación (denominadas facetas) (Cronbach, Gleser, Nanda, y Rajaratnam, 1972). El primer objetivo es analizar las diferentes fuentes de variación que pueden estar afectando a una medida o diseño de medida de origen observacional o correlacional. Su aplicación 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. En este caso el problema de la fiabilidad se refiere a la precisión de esta generalización. La TG ha contribuido a unificar las diferentes definiciones de fiabilidad, validez y precisión. Estas definiciones son complementarias sobre aspectos parciales de un modelo general, que considera todas las fuentes de variación que afectan a los resultados observados. Uno de los objetivos importantes de la medición es intentar identificar y medir los componentes de variancia que están aportando 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; Blanco-Villaseñor, Castellano, Hernández-Mendo, Sánchez-López, y Usabiaga, 2014).

El tercero de los objetivos del análisis llevado a cabo con esta teoría es desglosar, en cualquier tipo de medición, la variabilidad real de la variabilidad del error. El cumplimiento de este objetivo pasa ineludiblemente por la estimación de los componentes de variancia de diversas facetas que componen el estudio observacional. La cuestión central de la TG, por tanto, se encuentra en los componentes de variancia, dado que su magnitud aporta información sobre las fuentes de error que están afectando una medición conductual. El investigador definirá el conjunto de condiciones que restringen las facetas. El conjunto de valores que puede tomar un factor se le 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 las componentes de error; (4) Optimización de los coeficientes de Generalizabilidad.

La resolución de los problemas de cálculo que presenta la TG han motivado el desarrollo de diferentes programas informáticos, con características distintas. Algunos de estos programas limitan la potencia de cálculo. Actualmente se pueden encontrar distintos programas, a saber:

1. **Genova.** Desarrollado por Robert L. Brennan (Universidad de Iowa) usando el lenguaje de programación C++. Funciona

bajo los sistemas operativos: Windows, Mac, Linux/Unix. La licencia es gratuita. Aplicación de línea de comandos muy ligera y eficiente. Carga los datos a partir de un fichero de texto. El diseño debe ser balanceado u ortogonal. El máximo número de facetas es seis. No genera gráficos (Brennan, 1983, 2003).

2. **GT 2.0 E.** Desarrollado por Pierre Ysewijn. Funciona bajo sistema operativo MS-DOS. La licencia es gratuita. La aplicación cuenta con una interfaz gráfica pero presenta cuatro problemas: (1) Compatibilidad, no existe versiones para otras plataformas. (2) Los ficheros se almacenan en una ruta de acceso concreta y si un fichero no está en dicha ruta no puede ser abierto desde la aplicación. (3) Está limitado a ocho facetas. (4) El archivo de datos no puede sobrepasar el tamaño de 64 KB. (5) Solo permite realizar estudio de Generalizabilidad sobre facetas que tengan un solo nivel de anidamiento. (6) No se pueden imprimir los resultados ni exportar los datos a otras aplicaciones como Microsoft Office, Word o Excel (Ysewijn, 1996).
3. **EduG 6.1.** Desarrollado por *Swiss Society for Research in Education Working Group* en C++ para sistema operativo Windows. La licencia es gratuita. Es un proyecto de colaboración entre tres universidades de Canadá, Francia y Suiza. Es una aplicación con interfaz gráfica. No genera gráficos, pero permite emitir informes en formato txt y rtf. Limitado a ocho facetas. La versión 5.0 contenía errores de cálculo. La versión 6.0 corregía parte de estos errores. Sin embargo, se ha detectado que se generaban valores diferentes si se partía de una tabla de frecuencias o de la suma de los cuadrados de esta. Aunque la versión 6.1 corrige un error al salvar los datos en un nuevo fichero, el error anteriormente comentado persiste (Cardinet, Johnson, y Pini, 2010).

Existen otros paquetes estadísticos (SPSS, SAS 9.3, Matlab 7.14, Scilab 5.4.0) que pueden realizar estos cálculos, pero no de forma automática

La utilización de la TG, aunque no es amplia, es diversa, p.e. en análisis del juego (Blanco-Villaseñor, Castellano, y Hernández-Mendo, 2000), programas de actividad física (Morales-Sánchez, Blanco-Villaseñor, y Hernández-Mendo, 2004), análisis cualitativo de la violencia en el deporte (Molina-Macias y Hernández-Mendo, 2004), tensiomiografía (García-García, Hernández-Mendo, Serrano-Gómez, y Morales-Sánchez, 2013), entrenamiento deportivo (Casamichana, Castellano, y Hernández-Mendo, 2014) o en el perfil psicológico de deportistas (Álvarez et al., 2014).

El objetivo de este trabajo fue diseñar y construir un programa informático que realice los cálculos necesarios para un estudio de Generalizabilidad, superando las limitaciones presentadas por los programas existentes y mostrar su aplicabilidad en una investigación.

Tabla 1

Tabla comparativa de programas informáticos que realizan análisis de generalizabilidad.

Cuadro Resumen					
Logo/ Nombre	Licencia	Implementa la teoría	Ventajas	Inconvenientes	Realiza Gráficos
 Genova	Gratis	Sí	Es muy liviano	Solo 6 facetas	No
 GT 2.0 E	Gratis	Sí	Los resultados son fiables	Dificultades para compartir información. Limitado a 8 facetas.	No
 EduG 6.1	Gratis	Sí	Simple y fácil de manejar	Poco fiable. Limitado a 8 facetas.	No

Descripción del programa SAGT

SAGT¹ (acrónimo de *Software Application for Generalizability Theory*) es una herramienta multilinguaje (inglés, español, francés y portugués) para la aplicación de la Teoría de la Generalizabilidad (TG). Puede realizar tareas de estimación y optimización de las facetas de estudio de manera automática, incluyendo representaciones gráficas de las estimaciones realizadas.

Esta aplicación mejora las prestaciones de las reseñadas anteriormente, tiene más capacidades y es sencilla e intuitiva. Permite la importación y exportación de los datos desde y hacia otras aplicaciones.

Las disponibilidades de las opciones mostradas en el menú principal variarán en función del estado del programa (si se encuentra conectado a la plataforma MenPas www.menpas.com – Plataforma Psicosocial de Evaluación on-line) y del rol del usuario de la aplicación.

La primera opción del menú permite la conexión/desconexión con los servicios de la plataforma MenPas. La pantalla principal presenta un área de trabajo compuesta por pestañas, donde se muestran las tablas y los campos de texto y donde se realiza la edición de tablas y comentarios. Permite trabajar hasta con 16 facetas, e importa datos de GT, EduG

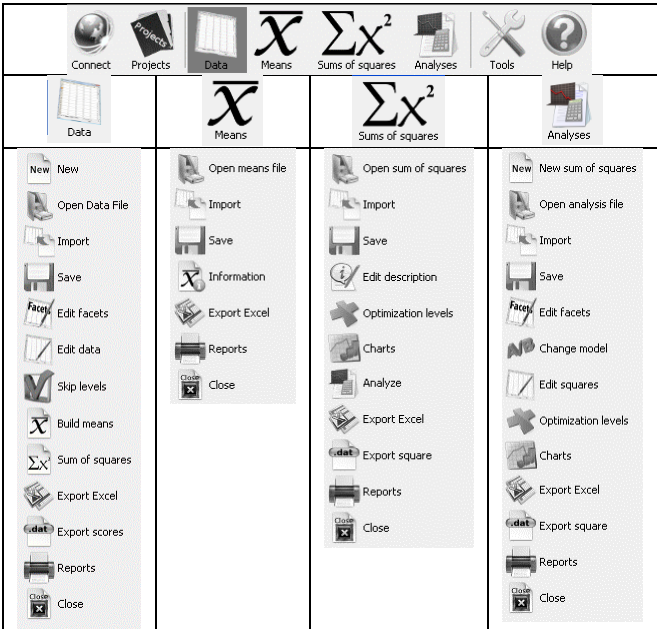


Figura 1. Menú Principal

¹ Es descargable como software libre desde la plataforma www.menpas.com

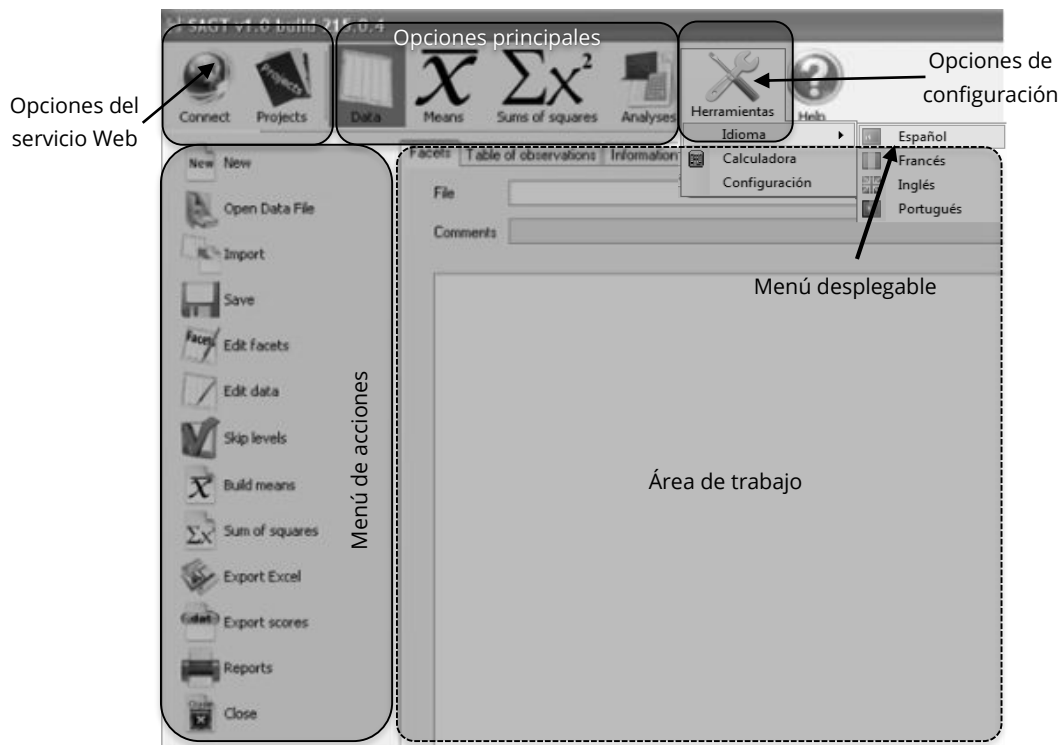


Figura 2a. Ventana principal de la Aplicación

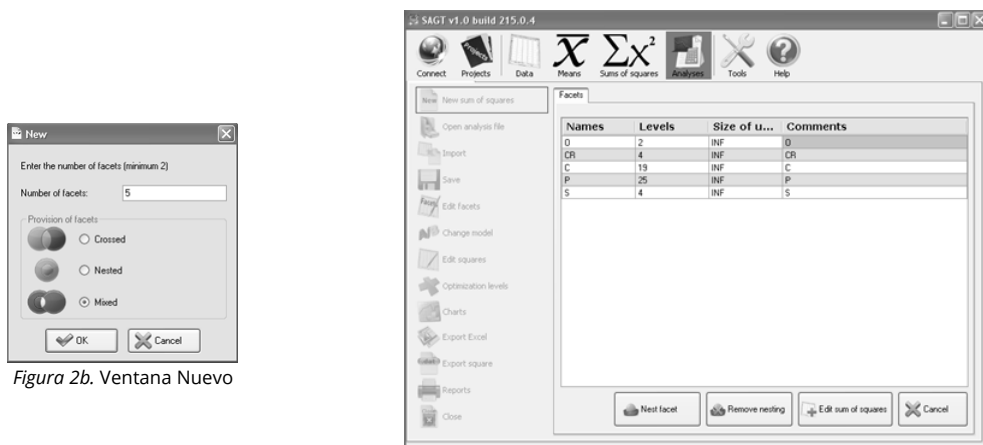


Figura 2b. Ventana Nuevo

Figura 3. Edición de facetas

SAGT permite importar puntuaciones de ficheros de texto compatibles con la mayoría de programas de cálculo estadístico como SPSS, SAS, Matlab, EduG 6.0, etc., siempre que estén en un fichero de texto que use el espacio o el retorno de carro como separador.

Se puede recuperar la información guardada por SAGT. Los archivos .sagt pueden contener datos de las facetas y tabla de frecuencias, tablas de medias y/o tablas de análisis. Los datos no son excluyentes entre sí. El programa recuperará toda la información con independencia de su contenido, mostrándola en la opción correspondiente. La aplicación cuenta con un directorio de trabajo por defecto donde se puede almacenar y recuperar los ficheros y si se dispone de acceso a la plataforma MenPas y los permisos necesarios se podrá recuperar archivos de datos almacenados en dicho servicio Web.

Permite modificar aquellos atributos de las facetas que no afecten al tamaño de la tabla de frecuencias. También permite editar tabla de frecuencias para modificar los valores de la variable observada en la tabla de frecuencias.

SAGT permite importar datos de ficheros generados por distintas aplicaciones para construir con ellos una tabla de frecuencias. Los tipos de archivos importables son los siguientes: (1) *Fichero de observaciones (.obs)* GT E 2.0 (Ysewijn, 1996); (2) *Medias Aritméticas (.rsm)* GT E 2.0 (Ysewijn, 1996); (3) *Tabla de frecuencias exportada a Excel (.xls)*; (4) *Ficheros (.csv)*.

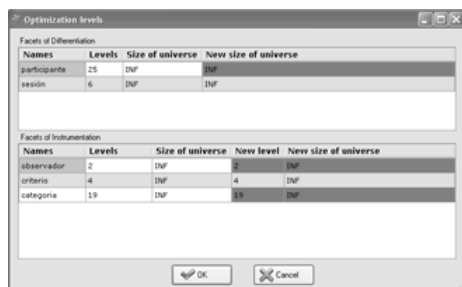


Figura 4. Ventana Niveles de optimización

La opción *Medias* permite su estudio a través de una tabla y permite importar tablas de medias aritméticas de las siguientes aplicaciones: (1) *Medias Aritméticas (.rsm)* GT E 2.0 (Ysewijn, 1996); (2) *Informe de medias (.rtf, .txt)* EduG 6.0 (Cardinet, Johnson, y Pini, 2010); (3) *Tablas de medias exportadas a Excel (.xls)* (ver Figura 1).

La opción *Suma de cuadrados* permite el estudio de análisis de las componentes de varianza, el cálculo de G-Parámetros, los coeficientes de Generalizabilidad y la estimación de dichos coeficientes con nuevos niveles para las facetas de estudio. Al igual que en opciones anteriores, se pueden recuperar los datos de un archivo almacenado localmente en el sistema y *Abrir desde servicio Web* para recuperar los datos compartidos en la plataforma Web. Igualmente dispone de dos opciones: *Guardar* y *Guardar en servicio Web*. Los tipos de archivos importables son los siguientes: (1) *Suma de cuadrados (.ssq)* GT E 2.0 (Ysewijn, 1996); (2) *Resultado de análisis de suma de cuadrados (.rsa)* GT E 2.0 (Ysewijn, 1996); (3) *Informe de media (.rtf, .txt)* (Cardinet, Johnson, y Pini, 2010); (4) *Tablas de medias exportadas a Excel (.xls)* (ver Figura 1 y 3).

SAGT permite la Optimización del diseño de medida y la estimación de G-Parámetros a partir de los nuevos niveles para las facetas de estudio. Se podrá modificar el tamaño del universo de todas las facetas y el nivel de las facetas de instrumentación. Los nuevos parámetros estimados se añadirán a la tabla resumen de G-Parámetros, como una nueva columna (Figuras 4 y 5).

Name of the values	Resume	Resume 2
observador	(2; INF)	(1; INF)
criterio	(4; INF)	(3; INF)
categoría	(19; INF)	(16; INF)
participante	(25; INF)	(25; INF)
sesión	(6; INF)	(6; INF)
Total number of observations	22800	7200
Coefficient G relative	0.984	0.970
Coefficient G absolute	0.946	0.914
Relative Error	0.004	0.007
Absolute Error	0.013	0.021
Stan. deviation of relative error	0.061	0.082
Stan. deviation of absolute error	0.112	0.144

Figura 5. Tabla resumen de optimización

SAGT permite realizar los siguientes tipos de gráficas:

- *Gráfico optimización* donde se muestran los datos de la tabla de resumen de datos.

Gráfico de Coef G rel. y Coef G abs (representa los valores del coeficiente G para las facetas de instrumentación seleccionadas). El gráfico mostrado puede ser representado mediante esplín (suavizado de curvas) o mediante representación lineal Figuras 6, 7 y 8).

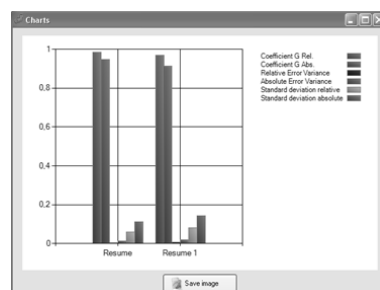


Figura 6. Gráfico de Barras coeficientes G, Error de la Varianza y de la desviación estándar

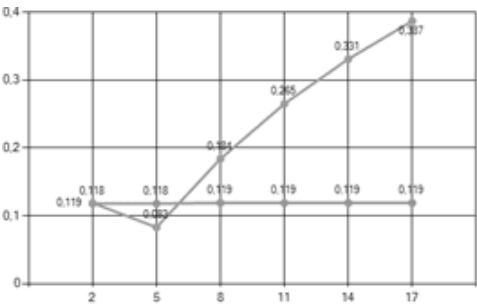


Figura 7. Ventana Gráfico de optimización

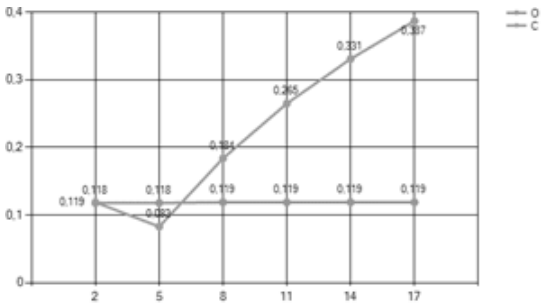


Figura 8. Coeficiente G absoluto: a la izquierda mediante esplín, a la derecha representación lineal

La opción *Análisis* permite el estudio de las tablas de sumas de cuadrados y de las componentes de varianza, el cálculo de G-Parámetros, los coeficientes de Generalizabilidad y la estimación de dichos coeficientes con nuevos niveles para las facetas de estudio. Al igual que en los casos anteriores se permite la importación /exportación de datos, así como la realización de gráficos e informes. El manual de usuario se puede descargar en www.menpas.com

A continuación, se presenta la aplicación de la Teoría de la Generalizabilidad a través del software SAGT, en una investigación sobre la existencia de discriminación de género en las clases de EF, siguiendo los trabajos de Álvarez-Bueno et al. (1990); Hastie y Pickwell (1996); Hellín, Moreno, y Rodríguez (2004); Ramos-Nateras y Hernández-Mendo, (2014); Sicilia (2003); Tannehill et al. (1994); Vázquez, Fernández-García, y Ferro (2000).

Tabla 2
Distribución de la muestra

	5ªA		5ªB		5ªC		5ªD		TOTAL
Muestra	25	%	25	%	25	%	25	%	100
Alumnas	14.00	27.45	12.00	23.53	13.00	25.49	12.00	23.53	51.00
Alumnos	11.00	22.45	13.00	26.53	12.00	24.49	13.00	26.53	49.00

Material

Las grabaciones se realizaron utilizando dos cámaras digitales marca SONY Handicam. El registro observacional se ha realizado utilizando, un sistema mixto de formatos de campo y sistema de categorías (ver tabla 3). Este sistema está constituido por 4 criterios y 13 categorías. Se ha utilizado el programa informático Hoisan 1.2 (Hernández-Mendo, López-López, Castellano, Morales-Sánchez, y Pastrana, 2012). Los análisis de generalizabilidad se realizaron mediante el programa SAGT (Hernández-Mendo, Ramos-Pérez, y Pastrana, 2012).

MÉTODO

Diseño

El diseño observacional utilizado está situado en el IV cuadrante, de carácter nomotético, seguimiento y multidimensional (N/S/M) (Anguera, Blanco-Villaseñor, Hernández-Mendo y Losada, 2011). Con respecto al diseño de generalizabilidad es ortogonal con 6 facetas (OBSERVADORES, CATEGORÍAS, CRITERIOS, PARTICIPANTE, CURSOS y SESIONES) y parcialmente anidado (dos facetas anidadas, categorías en criterio y participantes en cursos).

Participantes

La investigación se realizó en un Centro de Educación Primaria de la provincia de Málaga (España) con una muestra formada por 100 alumnos con edades comprendidas entre los 10 y los 12 años. Todos ellos cursaban 5º de Educación Primaria y se encontraban distribuidos en cuatro grupos clasificados como 5ªA, 5ªB, 5ªC y 5ªD. La distribución de la muestra por grupos y género se recoge en la Tabla 2.

Tabla 3
Herramienta de observación

CRITERIO 1. Conductas positivas hacia iguales de género opuesto.
1.1. Conducta de aceptación (ACEPT).
1.2. Conducta de relación (RELAC)
1.3. Conducta de ayuda y colaboración (ACOLAB)
CRITERIO 2. Conductas negativas hacia iguales de género opuesto.
2.1 Conducta de separación (SEPA)
2.2 Conducta rechazo (RECH)
2.3 Agresión verbal (VERB)
2.4 Agresión física (FÍSICA)
CRITERIO 3. Cooperación en la actividad
3.1 Conducta de cooperación completa (COOP)
3.2 Conducta de transformación de la tarea (TRAN)
3.3 Conducta de no cooperación (NO)
CRITERIO 4. Zonas de trabajo
4.1 Zonas principales (PRINC)
4.2 Zonas intermedias (INTER)
4.3. Zonas secundarias (MARGI)

Procedimiento

Las sesiones de E.F objeto de investigación fueron desarrolladas durante seis meses correspondientes al curso escolar 2012-2013, en el horario de E.F de cada grupo (dos horas semanales), impartidas por el profesor habitual.

Se recibieron firmados los correspondientes consentimientos informados para la grabación del alumnado por parte de los padres. Además, se obtuvo permiso de la dirección del centro respetándose, además, los principios de la declaración de Helsinki (revisión de 1983).

Se realizaron varias grabaciones piloto con el objetivo de (1) estimar errores de sesgo, de reactividad y reactancia; (2) decidir la ubicación definitiva de las videocámaras; (3) poner a prueba la herramienta observacional. Todas las grabaciones fueron realizadas en el gimnasio del centro.

La programación de EF utilizada durante esta investigación constaba de tres fases diferenciadas (ver tabla 3):

Tabla 4

Diseño del programa de EF utilizado para la investigación. Agrupamientos mixtos (X), Agrupamientos por género (NX), Agrupamientos por nivel de habilidad (H).

Tipo contenido	LÍNEA BASE PRE INTERVENCIÓN			INTERVENCIÓN							LÍNEA BASE POST INTERVENCIÓN				
	Masc	Alter	Fem								Masc	Alter	Fem		
Contenido sesiones	Fuerza	Velocidad	Resistencia	Pelota gigante	Hockey	Danzas mundo	Visionado videos género	Speed dating	Contenidos alternativos y Competiciones mixtas		Fuerza	Velocidad	Resistencia	Pelota gigante	Hockey
Nº Sesión	1ª 2ª	3ª	4ª	5ª-7ª	8ª	9ª-12ª	1ª	2ª	3ª-7ª		1ª 2ª	3ª	4ª	5ª-7ª	8ª
5ªA	L	H	NX	X	L	H	En clase	En clase	X		L	H	NX	X	L
5ªB	X	L	H	NX	X	L			X		X	L	H	NX	X
5ªC	NX	X	L	H	NX	X			X		NX	X	L	H	NX
5ªD	H	NX	X	L	H	NX			X		H	NX	X	L	H

Las sesiones de EF de las fases pre y post intervención se distribuyeron en tres bloques de contenidos en función de su estereotipo de género asociado: contenidos masculinos (fuerza, velocidad, etc.), contenidos femeninos (expresión corporal, baile, etc.) y contenidos alternativos o mixtos (pelota gigante, hockey, etc.). Las sesiones eran las mismas para todos los grupos pero en cada uno de ellos se fueron alternando cuatro tipos de agrupamientos diferentes: mixtos (X), donde debía haber chicos y chicas dentro de cada grupo; por género (NX), chicos y chicas se agrupaban según su género; por nivel de habilidad (H), donde los grupos se formaron según el grado de dominio que cada uno tenía de la actividad, independientemente del género; y libres (L), los agrupamientos se realizaban libremente sin la intervención del docente.

En la fase de intervención (7 sesiones) se llevaron a cabo tres técnicas de intervención: (1) *Speed dating*. Se realizó en el aula de clase con el propósito de fomentar la relación y el conocimiento mutuo entre los chicos y las chicas de cada grupo participante. Se organizaron los pupitres formando un doble círculo con las chicas colocadas en el interior y los chicos en el exterior. Cada pareja disponía de dos minutos para obtener la mayor información posible sobre su compañero/a. Cuando el tiempo se agotaba, el círculo exterior rotaba y los participantes cambiaban de pareja, iniciándose de nuevo la tarea. (2) *Visionado de videos*. Se proyectaron dos pequeños videos acerca de la problemática de la discriminación de género en la sociedad. Al finalizar el visionado de los mismos, se realizó un pequeño debate seguido de una actividad donde los alumnos debían identificar situaciones de discriminación de género en su ámbito familiar y escolar. (3) *Mural de la igualdad de género*. Se colocó un gran mural en la pared donde el alumnado podía anotar las

conductas de género surgidas durante la clase, ya fueran negativas (con bolígrafo rojo) o positivas (con bolígrafo azul). Al final de la clase, todos los alumnos se sentaban frente al mural y se procedía a la lectura, discusión y reflexión de las conductas anotadas. (4) *Leyendas deportivas*. Con el fin de dar a conocer al alumnado la existencia de grandes deportistas femeninas, durante la fase de intervención se dedicaron los primeros minutos de cada sesión a presentar mediante fotografías y videos a algunas de las más famosas deportistas femeninas de la historia (p.e., Nadia Comaneci, Yelena Isinbayeva, o Steffi Graf).

Las grabaciones fueron observadas por un único observador que recibió entrenamiento a lo largo de un mes durante tres sesiones semanales de una hora, al final del cual se realizó un análisis de calidad del dato.

RESULTADOS

Se realizó un análisis de calidad del dato de la herramienta observacional. El índice *kappa* de Cohen (1960) calculado con el programa informático Hoisan, obtiene unos valores para la concordancia intra-observadores de 0.81 considerando todos los participantes en conjunto, y una media 0.73 considerando cada uno de los participantes individualmente.

En el análisis *Generalizabilidad*, se utiliza un diseño de 6 facetas parcialmente anidado ([OBS][CAT]:[CRI] [PAR] [CUR] [SES]). Utilizando una de las características del programa SAGT se calculan modelos parciales. Para estimar la *homogeneidad*, la bondad de ajuste de las categorías y la fiabilidad de los observadores, se utilizaron tres diseños de tres facetas [OBS] [CRI] /[CAT], [OBS] /[CAT] [CRI] y [CAT] [CRI] /[OBS] (ver Tabla 5). Se obtuvieron índices G para los dos primeros modelos iguales

a 0, lo que indica que el sistema de observación ajusta y sus categorías son exhaustivas y mutuamente excluyentes (E/ME). En el tercer modelo los índices G son iguales a 1, esto indica una alta fiabilidad de los observadores. El mayor porcentaje de

varianza (99.9%), está asociado a las categorías anidadas en los criterios ([CAT]:[CRI]), siendo muy bajas para *observadores* (0%) y para las interacciones.

Tabla 5

Análisis de generalizabilidad modelos parciales ajuste de categorías fiabilidad observadores con SAGT.

Fuentes de variación	Suma de cuadrados	gl	Cuadrado medio	Aleatorio	Mixtos	Corregidos	%	Error estándar
[OBS]	35.779	1	35.779	0	0	0	0	0,724
[CAT]:[CRI]	251202886.6	48	5233393.471	2616678.846	2616678.846	2616678.846	99.999	523339,347
[OBS]:[CAT]:[CRI]	1717.385	48	35.779	35.779	35.779	35.779	0.001	7,156
[CRI]	11735467.72	3	3911822.574	-50829.65	-50829.65	-50829.65	0	103321,17
[OBS]:[CRI]	107.337	3	35.779	0	0	0	0	1,826
DISEÑOS								
Índices G	[OBS] [CRI] /[CAT]		[OBS] /[CAT] [CRI]		[CAT] [CRI] /[OBS]			
G relativo	0.0		0.0		1.0			
G absoluto	0.0		0.0		1.0			

También con modelos parciales se estimó la constancia intersesional e intergrupo, utilizando para ello cuatro facetas [OBS]:[PAR]:[CUR]:[SES], estando la faceta PARTICIPANTES anidado en la faceta CURSOS. Los índices G en los tres modelos estimados son óptimos dando cuenta de la estabilidad de ambas medidas. El mayor porcentaje de varianza está asociado [PAR]:[SES]:[CUR] (72.302%) (ver tabla 6).

En el análisis del modelo total, el mayor porcentaje de varianza está asociado a la interacción [CAT]:[PAR]:[SES]:[CRI]:[CUR] (61.668%). Los índices G para la estimación de la fiabilidad de los observadores ha sido 1. En la estimación del ajuste del sistema

de criterio y categorías ([OBS] [PAR] [CUR] [SES] /[CAT] [CRI]) los índices G fueron 0.079 para el relativo y 0.064 para el absoluto. Estos índices en la estimación de la faceta categorías es 0.025 y 0.020 respectivamente (ver tabla 6).

El plan de optimización se ha realizado numérica y gráficamente, como se puede comprobar en la Tabla 7. Se ha realizado una estimación del número de sesiones, cursos y participantes necesarios para alcanzar índices G óptimos. Con la muestra actual se obtiene que ambos índices G se sitúan en 0.984.

Tabla 6

Análisis de generalizabilidad de modelos parciales, calculando la constancia intersesional e intergrupo con SAGT

Fuentes de variación	Suma de cuadrados	gl	Cuadrado medio	Aleatorio	Mixtos	Corregidos	%	Error estándar
[OBS]	3.101	1	3.101	0.005	0.005	0.005	0	0.004
[PAR]:[CUR]	688381.893	96	7170.645	406.125	406.125	406.125	25.574	86.251
[OBS]:[PAR]:[CUR]	76.053	96	0.792	0.049	0.049	0.049	0.003	0.02
[CUR]	15143.656	3	5047.885	-12.603	-12.603	-12.603	0	12.067
[OBS]:[CUR]	0.929	3	0.31	-0.003	-0.003	-0.003	0	0.002
[SES]	10099.004	5	2019.801	-9.678	-9.678	-9.678	0	8.701
[OBS]:[SES]	2.444	5	0.489	0	0	0	0	0.003
[PAR]:[SES]:[CUR]	1102490.347	480	2296.855	1148.177	1148.177	1148.177	72.302	73.977
[OBS]:[PAR]:[SES]:[CUR]	240.587	480	0.501	0.501	0.501	0.501	0.032	0.032
[CUR]:[SES]	59332.399	15	3955.493	33.173	33.173	33.173	2.089	27.295
[OBS]:[CUR]:[SES]	7.386	15	0.492	0	0	0	0	0.007
DISEÑOS								
Índices G	[OBS] [PAR] [CUR] [SES]		[OBS] [PAR] [SES] /[CUR]		[OBS] [PAR] [CUR] /[SES]			
G relativo	0.895		0.995		0.673			
G absoluto	0.892		0.995		0.673			

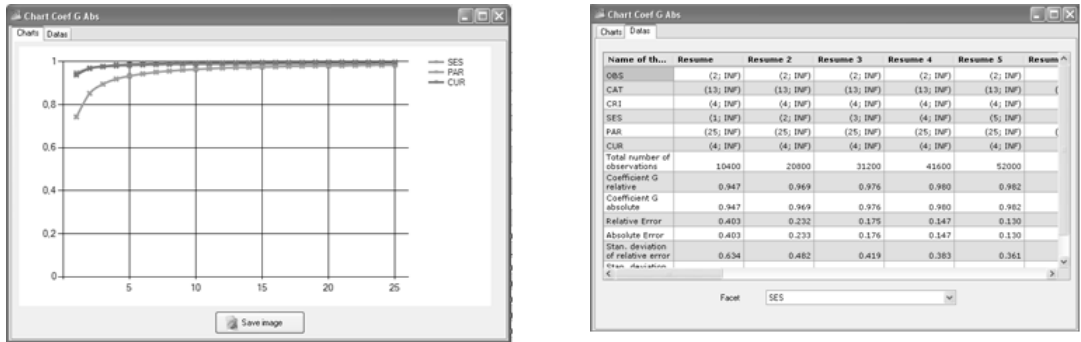


Figura 9. Optimización gráfica y numérica del diseño de medida

Se realizó una comparación de medias (ANOVA) en relación al sexo de los participantes y la conducta objeto de estudio. Éstos indican, en la fase de pretest, la existencia de diferencias significativas entre chicos y chicas en cuanto a la autoría de las

conductas de discriminación por sexo en “conductas de agresión verbal y física” ($F = 4.479, p = .046$); “conductas de separación y rechazo” ($F = 4.800, p = .049$)

Tabla 7

Análisis de generalizabilidad de modelos total con SAGT.

Fuentes de variación	Suma de cuadrados	gl	Cuadrado medio	Aleatorio	Mixtos	Corregidos	%	Error estándar
[OBS]	0.06	1	0.06	0	0	0	0	0
[CAT]:[CRI]	418671.478	48	8722.322	7.207	7.207	7.207	19.941	1.454
[OBS]:[CAT]:[CRI]	2.862	48	0.06	0	0	0	0	0
[CRI]	19559.113	3	6519.704	-0.142	-0.142	-0.142	0	0.287
[OBS]:[CRI]	0.179	3	0.06	0	0	0	0	0
[PAR]:[CUR]	13238.113	96	137.897	0.043	0.043	0.043	0.12	0.036
[OBS]:[PAR]:[CUR]	1.463	96	0.015	0	0	0	0	0
[CAT]:[PAR]:[CRI]:[CUR]	543815.028	4608	118.015	6.119	6.119	6.119	16.931	0.208
[OBS]:[CAT]:[PAR]:[CRI]:[CUR]	70.203	4608	0.015	0.001	0.001	0.001	0.003	0
[CRI]:[PAR]:[CUR]	32079.806	288	111.388	-0.043	-0.043	-0.043	0	0.062
[OBS]:[CRI]:[PAR]:[CUR]	4.388	288	0.015	0	0	0	0	0
[CUR]	291.224	3	97.075	-0.003	-0.003	-0.003	0	0.005
[OBS]:[CUR]	0.018	3	0.006	0	0	0	0	0
[CAT]:[CUR]:[CRI]	14064.648	144	97.671	-0.144	-0.144	-0.144	0	0.041
[OBS]:[CAT]:[CUR]:[CRI]	0.858	144	0.006	0	0	0	0	0
[CRI]:[CUR]	880.83	9	97.87	0.003	0.003	0.003	0.007	0.012
[OBS]:[CRI]:[CUR]	0.054	9	0.006	0	0	0	0	0
[SES]	194.212	5	38.842	-0.002	-0.002	-0.002	0	0.004
[OBS]:[SES]	0.047	5	0.009	0	0	0	0	0
[CAT]:[SES]:[CRI]	10634.859	240	44.312	-0.116	-0.116	-0.116	0	0.027
[OBS]:[CAT]:[SES]:[CRI]	2.256	240	0.009	0	0	0	0	0
[CRI]:[SES]	692.027	15	46.135	0.002	0.002	0.002	0.005	0.008
[OBS]:[CRI]:[SES]	0.141	15	0.009	0	0	0	0	0
[PAR]:[SES]:[CUR]	21201.737	480	44.17	-0.005	-0.005	-0.005	0	0.032
[OBS]:[PAR]:[SES]:[CUR]	4.627	480	0.01	0	0	0	0	0
[CAT]:[PAR]:[SES]:[CRI]:[CUR]	1027191.52	23040	44.583	22.287	22.287	22.287	61.668	0.208
[OBS]:[CAT]:[PAR]:[SES]:[CRI]:[CUR]	222.08	23040	0.01	0.01	0.01	0.01	0.027	0
[CRI]:[PAR]:[SES]:[CUR]	64397.556	1440	44.721	0.005	0.005	0.005	0.015	0.066
[OBS]:[CRI]:[PAR]:[SES]:[CUR]	13.88	1440	0.01	0	0	0	0	0
[CUR]:[SES]	1141.008	15	76.067	0.005	0.005	0.005	0.013	0.011
[OBS]:[CUR]:[SES]	0.142	15	0.009	0	0	0	0	0
[CAT]:[CUR]:[SES]:[CRI]	48631.544	720	67.544	0.459	0.459	0.459	1.271	0.072
[OBS]:[CAT]:[CUR]:[SES]:[CRI]	6.818	720	0.009	0	0	0	0	0
[CRI]:[CUR]:[SES]	2911.621	45	64.703	-0.005	-0.005	-0.005	0	0.021
[OBS]:[CRI]:[CUR]:[SES]	0.426	45	0.009	0	0	0	0	0

Tabla 8
Análisis descriptivos y ANOVA correspondientes a la recategorización de las categorías 2.1 (separación), 2.2 (rechazo), 2.3 (agresión verbal) y 2.4 (agresión física).

		Suma de cuadrados		gl	Media cuadrática		F	Sig.
Inter-grupos		2599100.167		1	2599100.167		8.124	.029
Intra-grupos		1919559.333		6	319926.556			
Total		4518659.500		7				
	N	M	DT	Error típico	Intervalo de confianza para la media al 95%		Mínimo	Máximo
					Límite inferior	Límite superior		
Pre	6	1349.83	619.577	252.941	699.63	2000.04	380	1971
Post	2	33.50	13.435	9.500	-87.21	154.21	24	43
Total	8	1020.75	803.444	284.060	349.05	1692.45	24	1971

Las conductas negativas por razón de sexo son significativamente mayores en las clases de EF previas a la intervención ($F = 0.29, p < .05$), reduciéndose drásticamente después de la misma. Se han obtenido resultados significativos también en la recategorización de las conductas positivas por razón de sexo entre las dos fases ($F = 0.41, p < .05$).

DISCUSIÓN

La utilización del programa SAGT ha resultado eficaz en el tratamiento y análisis de los datos a diversos niveles, especialmente en la gestión de opciones novedosas del programa, a saber:

- (1) La exportación de datos desde el programa de codificación HOISAN se ha realizado de forma automática a pesar de ser una matriz de 62400 líneas. En los programas existentes era necesario introducir manualmente o las frecuencias o la suma de cuadrados (vid GENOVA, GT ó EduG). Esto supone un ahorro de tiempo y un incremento de la fiabilidad al reducir los errores de tecleo.
- (2) Ha permitido, dentro del modelo global de 6 facetas parcialmente anidado ([OBS][CAT]:[CRI] [PAR] [CUR]:[SES]) utilizar la estimación de diversos modelos parciales que permiten aislar el peso de determinadas facetas, y determinar, a través de los índices G, que el sistema de observación ajusta y sus categorías son exhaustivas y mutuamente excluyentes (E/ME). También, que los índices G para los observadores indica una alta fiabilidad de los observadores. Se ha comprobado que el mayor porcentaje de varianza (99.9%), está asociado a las categorías anidadas en los criterios ([CAT]:[CRI]), siendo muy bajas para observadores (0%) y para las interacciones. Esto se ha podido comprobar tanto en los modelos parciales como en los modelos totales. Esta opción, al igual que otras que presenta el programa, era totalmente inédita en programas anteriores.
- (3) Ha permitido realizar los cálculos de la tabla de frecuencias utilizando la opción "Ocultar Nulos" (ver figura 15). Como ya ha quedado señalado, esta opción es novedosa y permite acelerar los cálculos.
- (4) Ha permitido la optimización del Plan de Medida a nivel gráfico y numérico. Esta última opción permite un ahorro de tiempo considerable en la optimización del diseño medida pues ofrece conjuntamente la representación gráfica y la estimación numérica del intervalo y puntos considerados. En esta investigación la estimación realizada sobre el número de sesiones, cursos y participantes necesarios para alcanzar índices

G óptimos, se ha obtenido que la muestra actual obtiene que ambos índices G se sitúan en 0.984 (ver figuras 8 y 9, tabla 7).

(5) Posibilita el trabajo hasta con 16 facetas y diversos niveles de anidamiento. El número de facetas y anidamientos y la estructura de los mismos con las que trabajan otros programas es sensiblemente inferior.

(6) Permite almacenar los datos en la plataforma www.menpas.com y compartirlos con otros investigadores. Todas estas características, junto a las ya descritas, permiten asegurar que el programa SAGT supera las limitaciones de los programas antecesores, siendo más eficaz y amigable en el tratamiento de los datos.

La eficacia del programa de intervención desarrollado para reducir la desigualdad de género entre el alumnado ha quedado constatada con los resultados estimados. Estos resultados son coherentes con los trabajos de Dunbar y O'Sullivan (1986) y Andersen y Wold (1992), y plantean la necesidad de desarrollar programas de igualdad entre géneros. También se sitúan en la línea de trabajos como los de Planchuelo (2008); Hernández-Mendo y Planchuelo (2012); Hernández-Mendo, Olmedo-Rueda, y Planchuelo (2012), donde se demostró la eficacia de intervenciones para el desarrollo moral. Estos resultados avalan la función del docente como mediador de aprendizaje y no sólo como mero transmisor de conocimientos (Aicinema, 1991; Cervelló et al., 2003, 2004; Gutiérrez-Sanmartín y Vivó, 2003; Harrison y Blakemore, 1989).

En relación al género del alumnado, los resultados indican que las conductas negativas de género son más frecuentes en los chicos que en las chicas. Jiménez (2004) considera que los chicos son más indisciplinados que las chicas. Los resultados también muestran que las agresiones verbales y físicas relacionadas con el género se producen de forma muy esporádica. No obstante, cuando éstas tienen lugar su autoría es mayoritariamente de los chicos (Derry, 2002).

No se ha encontrado ninguna relación entre las conductas positivas o negativas de género realizadas por el alumnado y el tipo de contenido masculino, femenino o alternativo utilizado. Riemer y Visio (2003) constatan una evolución en la percepción del conjunto de actividades "neutrales" sin adscripción de género. También encontraron que el grado de cooperación de los escolares es mucho mayor en las sesiones con contenidos alternativos que en las sesiones con contenidos vinculados al género masculino o femenino. Holland y Andre (1994)

argumentan que aquéllos(as) que se implican en actividades físicas más propias de su género son valorados mejor socialmente que aquellos que practican actividades relacionadas con el estereotipo del género opuesto. Taberner y Márquez (1997) encontraron que las chicas se integran mejor en las actividades habituales de los chicos que al contrario. Los

resultados obtenidos se sitúan en esta línea, las conductas de no cooperación han sido mucho más frecuentes en las sesiones de baile que en las de fuerza y velocidad. Terrón (2001), considera que este hecho podría estar en la gestión inadecuada de sentimientos y emociones del género masculino.

The screenshot shows the SAGT v1.0 build 215.0.4 application window. The 'Table of observations' tab is active, displaying a frequency table with the 'Hide null' checkbox checked. The table has columns for [OBS], [CAT]:[...], [CRI], [PAR]:[...], [CUR], [SES], and Measurement variable. The data is organized into rows representing different observations.

[OBS]	[CAT]:[...]	[CRI]	[PAR]:[...]	[CUR]	[SES]	Measurement variable
1	1	1	1	1	1	4
1	1	1	1	1	2	1
1	1	1	1	1	3	4
1	1	1	1	1	4	1
1	1	1	1	1	5	20
1	1	1	1	1	6	1
1	1	1	1	2	1	3
1	1	1	1	2	2	1
1	1	1	1	2	3	8
1	1	1	1	2	4	8
1	1	1	1	2	5	20
1	1	1	1	2	6	58
1	1	1	1	3	1	62
1	1	1	1	3	2	45
1	1	1	1	3	3	60
1	1	1	1	3	4	8
1	1	1	1	3	5	8
1	1	1	1	3	6	20

Figura 15. Tabla de frecuencias con opción de "Ocultar nulos"

El tipo de agrupamiento utilizado se ha mostrado como un factor determinante de las conductas de género entre el alumnado. Los agrupamientos mixtos y por nivel de habilidad son los más aptos para favorecer las conductas positivas de género, y el nivel de habilidad como el más idóneo para reducir la aparición de conductas negativas de rechazo (incluso más que los mixtos). Estos resultados sugieren que las conductas negativas de género estarían relacionadas con la participación en un grupo con un nivel de habilidad motriz desigual más que con la participación en un grupo mixto. Estos resultados confirman los obtenidos por Camarero, Tella, y Mundina (1997). Griffin (1984, 1985a, 1985b), encontró que algunas chicas son buenas oponentes para los adversarios masculinos, siendo reconocidas por éstos como buenas jugadoras y sintiéndose éstas cómodas cuando participan con los chicos. Otros estudios (Lirgg, 1993; Treanor et al, 1998) corroboran estos resultados y defienden que las alumnas más hábiles prefieren las clases mixtas frente a las menos hábiles, que se encuentran más seguras en clases del mismo género. En el polo opuesto se situarían los agrupamientos libres y por género, cuyos resultados han sido muy similares. Se ha constatado que cuando el profesor no interviene, los estudiantes se organizan en función del género. Ambos tipos de agrupamientos se caracterizan por la inexistencia de conductas positivas o de rechazo.

REFERENCIAS

Aicinema, S. (1991). The teacher and student attitudes toward physical education, *Physical Educator*, 48(1), 28-32.

Álvarez, O., Estevan, I., Falcó, C., Hernández-Mendo, A., y Castillo, I. (2014). Perfil de habilidades psicológicas en taekwondistas universitarios y su relación con el éxito en competición. *Cuadernos de Psicología del Deporte*, 14(3), 13-20.

Álvarez-Bueno, G., Cano, S., Fernández-García, E., López, C., Manzano, A., y Vázquez, B. (1990). *Guía para una educación física no sexista*. Madrid: Secretaría de Estado para la Educación, Ministerio de Educación y Ciencia.

Anderssen, N., y Wold, B. (1992). Parental and peer influences on leisure-time physical activity in young adolescents. *Research Quarterly for Exercise and Sport*, 63(4), 341-348.

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., Castellano, J., y Hernández-Mendo, A. (2000). Generalizabilidad de las observaciones en la acción del juego en el fútbol. *Psicothema*, 12(Supl. 2), 81-86.

Blanco-Villaseñor, A., Castellano, J., Hernández-Mendo, A., Sánchez-López, C. R., y Usabiaga, O. (2014). Aplicación de la TG en el deporte para el estudio de la fiabilidad, validez y estimación de la muestra. *Revista de Psicología del Deporte*, 23(1), 131-137.

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.

Brennan, R. L. (1983). *Elements of Generalizability Theory*. Iowa City, IO: American College Testing Program.

- Brennan, R. L. (2003). Coefficients and Indices in Generalizability Theory. *Center for Advanced Studies in Measurement and Assessment*, 1, 2-43.
- Camarero, S., Tella, V., y Mundina, J. J. (1997). *La actividad deportiva en el ámbito escolar*. Valencia: Promolibro.
- Cardinet, J., Johnson, S., y Pini, G. (2010). *Applying Generalizability Theory using EduG*. New York: Routledge.
- Casamichana, D., Castellano, J., y Hernández-Mendo, A. (2014). La Teoría de la Generalizabilidad aplicada al estudio del perfil físico durante juegos reducidos con diferente orientación del espacio en fútbol. *RICYDE Revista Internacional de Ciencias del Deporte*, 10(37), 195-205. doi: <http://dx.doi.org/10.5232/ricyde2014.03702>
- Cervelló, E., Del Villar, F., Jiménez, R., Ramos, L., y Blázquez, F. (2003). Clima motivacional en el aula, criterios de éxito de los discentes y percepción de igualdad de trato en función del género en las clases de Educación Física. *Enseñanza: Anuario Interuniversitario de Educación*, 21, 379-395.
- Cervelló, E., Jiménez, R., Del Villar, F., Ramos, L., y Santos-Rosa, F. J. (2004). Goal orientations, motivational climate, equality, and discipline of spanish physical education students. *Perceptual and Motor Skills*, 99, 271-283.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37-46.
- Cronbach, L. J., Gleser, G. C., Nanda, H., y Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York: John Wiley and Sons.
- Derry, J.A. (2002). Single-sex and coeducation physical education: Perspective of adolescent girls and female physical education teachers. *Melpomene Journal*, 22, 17-28.
- Dunbar, R. R., y O'Sullivan, M. M. (1986). Effects of intervention on differential treatment of boys and girls in elementary physical education lessons. *Journal of Teaching in Physical Education*, 5, 166-175.
- García García, O., Hernández Mendo, A., Serrano Gómez, 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.
- Griffin, P. S. (1984). Girls' participation patterns in a middle school team sports unit. *Journal of Teaching in Physical Education*, 4, 30-38.
- Griffin, P. S. (1985a). Boys' participation styles in a middle school physical education team sports unit. *Journal of Teaching in Physical Education*, 4, 100-110.
- Griffin, P. S. (1985b). Girls' and boys' participation styles in middle school Physical Education team sport classes: A description and practical applications. *Physical Educator*, 42, 3-8.
- Gutiérrez-Sanmartín, M., y Vivó, P. (2003). Desarrollo Sociomoral en Educación Física. En M. J. Mosquera, V. Gambau, R. Sánchez, y X. Pujadas. (Eds.), *Deporte y Postmodernidad* (pp. 471-478). Madrid: Librerías Esteban Sanz.
- Harrison, J., y Blakemore, C. (1989). *Instructional strategies for secondary school physical education*. Dubuque, IA: William C. Brown Publishers.
- Hastie, P. A., y Pickwell, A. (1996). Take your partners: A description of a student social system in a secondary school dance class. *Journal of Teaching in Physical Education*, 15, 171-187.
- Hellín, P., Moreno, J. A., y Rodríguez, P. L. (2004). Motivos de práctica físicodeportiva en la Región de Murcia. *Cuadernos de Psicología del Deporte*, 1 y 2(4), 101-116.
- Hernández-Mendo, A., Ramos-Pérez, F., y Pastrana, J.L. (2012). *SAGT: Programa informático para análisis de Teoría de la Generalizabilidad*. SAFE CREATIVE Código: 1204191501059.
- 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.
- Hernández-Mendo, A., Olmedo-Rueda, L., y Planchuelo, L. (2012). Cuestionario de Desarrollo Moral para Niños: Estudio Preliminar. *Avances de la Psicología del Deporte en Iberoamerica*, 1(1), 57-73.
- Hernández-Mendo, A., López-López, J. A., Castellano, J., Morales-Sánchez, V., y Pastrana, J. L. (2012). Hoisan 1.2: Programa informático para uso en metodología observacional. *Cuadernos de Psicología del Deporte*, 12(1), 55-78.
- Holland, A., y Andre, T. (1994). Athletic participation and the social status of adolescent males and females. *Youth and Society*, 25(3), 388-407.
- Jiménez R. (2004). *Motivación, trato de igualdad, comportamientos de disciplina y estilos de vida saludables en estudiantes de Educación Física en Secundaria*. Tesis Doctoral sin publicar. Universidad de Extremadura, Cáceres, España.
- Lirgg, C. D. (1993). Effects of same-sex versus coeducational physical education on the self-perceptions of middle and high school students. *Research Quarterly for Exercise of Sport*, 64, 324-334.
- Martínez-Arias, M. R. (1995). *Psicometría: teoría de los tests psicológicos y educativos*. Madrid: Síntesis.
- Molina-Macias, M. I., y Hernández-Mendo, A. (2004). La violencia en el deporte través de la prensa escrita. *Metodología de las Ciencias del Comportamiento, volumen especial*, 423-429.
- Morales-Sánchez, V., Blanco-Villaseñor, A., y Hernández-Mendo, A. (2004). Optimización de modelos de medida en la evaluación de programas de actividad física. *Metodología de las Ciencias del Comportamiento, volumen especial*, 437-443.
- Planchuelo, L. (2008). *Intervención y evaluación del desarrollo moral en educación física en primaria*. Tesis Doctoral sin publicar. Universidad de Málaga, Málaga, España.
- Ramos-Nateras, F., y Hernández-Mendo, A. (2014). Intervención para la reducción de la discriminación por sexo en las clases de educación física según los contenidos y agrupamientos utilizados. *Revista Española de Educación Física y Deportes*, 404, 27-38.
- Riemer, B. A., y Visio, M. E. (2003). Gender Typing of Sports: An Investigation of Metheny's Classification. *Research Quarterly for Exercise and Sport*, 7(2), 193-204.
- Sicilia, A. (2003). La investigación sobre el pensamiento del alumnado. Una revisión desde la Educación Física. *Revista de Educación*, 331, 577-613.
- Taberner, B., y Márquez, S. (1997). Análisis de las relaciones chico-chica en las clases de Educación Física. Un ejemplo práctico de investigación-acción en la formación inicial del

- profesorado de Educación Física. *Revista Española de Educación Física y Deportes*, 1, 34-39.
- Tannehill, D., Romar, J., O'Sullivan, M., England, K., y Rosenberg, D. (1994). Attitudes toward Physical Education: Their impact on how Physical Education teachers make sense of their work. *Journal of Teaching in Physical Education*, 13, 323-441.
- Terrón, E. (2001). Coeducación y control social en la España de la posguerra. *Revista de Educación*, 326, 185-193.
- Treanor, L., Graber, K., Housner, L., y Wiegand, R. (1998). Middle School Students' Perceptions of Coeducational and Same-Sex Physical Education Classes. *Journal of Teaching in Physical Education*, 18, 43-56.
- Vázquez, B., Fernández-García, E., y Ferro, S. (2000). *Educación Física y género: modelos para la observación y el análisis del comportamiento del alumnado y del profesorado*. Madrid: Gymnos.
- Ysewijn, P. (1996). *Generalizability Study. Versión 2.0.E*. Mimeografía.

Agradecimientos

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.