

Diseño y validación de un instrumento de observación de las conductas que alteran la convivencia en educación física

Design and Validation of an Observational Instrument of Student Misconducts in Physical Education

**BERNARDINO J. SÁNCHEZ-ALCARAZ
MARTÍNEZ**

Universidad de Murcia
bjavier.sanchez@um.es

ALBERTO GÓMEZ-MÁRMOL

Universidad de Murcia
alberto.gomez1@um.es

ALFONSO VALERO-VALENZUELA

Universidad de Murcia
avalero@um.es

ROCÍO ESTEBAN LUIS

Universidad de Murcia
rocionkt@hotmail.com

SIXTO GONZÁLEZ-VÍLLORA

Universidad de Castilla-La Mancha
Sixto.gonzalez@uclm.es

Resumen: El objetivo del presente trabajo fue diseñar y validar un instrumento de observación de conductas que alteran la convivencia en las clases de Educación Física, que permita definir las, categorizarlas y medirlas cuantitativamente. Este proceso de validación se llevó a cabo en varias fases y contó con la participación de seis jueces expertos en el ámbito de la enseñanza de la Educación Física y cuatro observadores formados en el análisis de las conductas que alteran el clima de convivencia en las clases de Educación Física. Observaron un total de diez clases de Educación Física, de sexto de Educación Primaria y de tercero de la Educación Secundaria de diez centros escolares diferentes, analizando la estructura del instrumento mediante procedimientos confirmatorios. El instrumento final estuvo

formado por ocho conductas que alteran la convivencia en las clases de Educación Física, agrupadas en dos factores (conductas violentas y conductas de indisciplina y desinterés). Los resultados obtenidos en el proceso de validez, fiabilidad y consistencia interna indican que el instrumento diseñado es válido para registrar las conductas que alteran la convivencia en las clases de Educación Física y permite su aplicación sin la presencia del observador en el aula, minimizando así el efecto de esa presencia en el comportamiento del alumnado.

Palabras clave: Población en edad escolar, Instrumento de medida, Educación Física, Evaluación de la educación.

Abstract: The aim of this study was to design and validate an instrument of observation that would allow us to measure and quantify student misconduct during classes of physical education. This process of validation was carried out in several phases. Six expert judges in physical education teaching and four observers trained to measure student misconduct in physical education lessons participated in the process, analysing ten physical education lessons of sixth and ninth grade from ten different schools, settings the structure of the instrument through confirmatory procedures. The final instrument consisted of eight items describing different kinds of student misconduct during phys-

ical education classes, grouped into two categories (violent behaviours and undisciplined and disinterested behaviours). The results obtained in the process of validation, reliability and internal consistency suggest that the instrument designed is optimally useful and reliable to measure student misconduct during physical education classes, allowing its application without the presence of the investigator in the classroom, thus minimizing its effect on the behaviour of the student body.

Keywords: School age population, Measuring instruments, Physical Education, Educational evaluation.

INTRODUCCIÓN

Durante la adolescencia se experimentan, además de un rápido crecimiento físico, importantes modificaciones de índole conductual debidas a importantes cambios psico-sociales ligados al desarrollo (Escartí, Pascual y Gutiérrez, 2005). Estos cambios psicosociales comienzan a acentuarse al inicio de la adolescencia, entre los 10 y 12 años, y suponen en muchas ocasiones una crisis de identidad y valores, así como un incremento de conductas que alteran la convivencia, lo que se traduce con demasiada frecuencia en un considerable incremento de conflictos familiares y/o educativos (Borràs, Palou y Ponseti, 2001).

La convivencia educativa debe ser entendida como una construcción colectiva y dinámica, compuesta a partir del conjunto de interrelaciones humanas entre los agentes de la comunidad educativa, cuyos límites se definen por derechos y deberes, y cuya influencia trasciende el mero contexto escolar (García, 2010). Benítez, Tomás y Justicia (2005) entienden que la mejora de las habilidades sociales puede ayudar a la mejora de las relaciones interpersonales, estas relaciones pueden mejorar el clima de centro y, finalmente, la convivencia escolar. De modo que Peñalva, López-Goñi y Landa (2013) indican que el docente debe estar adecuadamente formado en competencias emocionales para proporcionar a sus alumnos una educación que les permita desarrollar dichas habilidades, contribuyendo así a un mejor clima de aula.

En los últimos años se vive en las aulas una crisis de valores y de convivencia, que se ve traducida a un aumento del absentismo escolar, acompañado de un clima de relación social en el que las intimidaciones, el maltrato, la xenofobia, la violencia o las drogas empiezan a ser muy comunes en la mayoría de los centros escolares (Ruiz, Rodríguez, Martinek, Schilling, Durán y Jiménez, 2006).

La participación de los estudiantes en comportamientos agresivos se ha asociado con resultados negativos, incluyendo fracaso escolar, inadaptación social, rechazo o comportamientos disruptivos y/o destructivos (Barnes, Smith y Miller, 2014). En este sentido, las conductas disruptivas en el entorno escolar, como la hostilidad o la agresión, pueden poner en peligro el mantenimiento de ambientes de aprendizaje óptimos en las escuelas (Kuhn, Ebert, Gracey, Chapman y Epstein, 2015).

Los cambios propios de la adolescencia pueden ser el origen de conflictos en el entorno educativo y, debido a la incidencia de conductas que alteran la convivencia, han proliferado en los últimos años estudios con programas de intervención orientados a la educación en materia de comportamiento socialmente adecuado. En estos programas se incluyen actividades físicas y/o deportivas, por ejemplo, en fútbol de formación, con la intención de inculcar valores de juego limpio (Palou, Borràs, Ponseti, Vidal y Torregrosa, 2007), o bien con un desarrollo positivo para las comunidades de jóvenes que están en algún grado de marginación (Whitley, Forneris y Barker, 2015).

En estudios recientes, también se han evaluado experiencias novedosas desde la perspectiva del formador, poniendo el objetivo de análisis en los profesores o su metodología. En España, los investigadores Fernández y Barreiras (2015) mejoraron la motivación y los comportamientos de los estudiantes de Educación Primaria a través del uso de las nuevas tecnologías, basadas en la animación y la creación audiovisual. Por otro lado, estudios internacionales mejoraron los comportamientos y la responsabilidad de estudiantes de Educación Secundaria a través de programas basados en la aplicación de estrategias de resolución de conflictos (Hemphill, Templin y Wright, 2015).

La mayor parte de los trabajos que estudian estos fenómenos se han realizado utilizando diferentes escalas y cuestionarios, o bien otro tipo de métodos cualitativos, como entrevistas (Urbina, Simón y Echeita, 2011). Sin embargo, son escasas las investigaciones que han utilizado metodologías observacionales para conocer los efectos de los diferentes programas para la mejora de los valores a través del deporte (Anguera, Blanco, Hernández y Losada, 2011). Una de las ventajas que ofrece este tipo de metodología es que permite el registro organizado y la cuantificación de las conductas perceptibles que se quieran analizar si se adecua un instrumento válido (Anguera et al., 2011). Dentro del contexto de convivencia escolar y, más concretamente, en el de conductas que alteran la convivencia en las clases de Educación Física, no existen apenas instrumentos que permitan registrar las conductas perceptibles.

En este sentido, uno de los primeros instrumentos de observación, basado en el sistema de categorías, diseñado para evaluar conductas que alteran la conviven-

cia y específico para el ámbito de la Educación Física, fue el *System of Observation of Disciplinary Incidents*, o bien SODI (Sistema de Observación de Incidentes de Disciplina), validado por Goyette, Dore y Dion (2000), que contiene 19 tipos de conductas inadecuadas, agrupadas en tres categorías según la intensidad (leve, grave o muy grave). Este sistema de categorías forma parte de un estudio centrado en estudiantes universitarios y futuros profesores, con la finalidad de descubrir y comprender las diferentes reacciones que los distintos profesores pueden aplicar en función del contexto en el que se manifiestan los comportamientos de los alumnos, por lo que no es válida para registrar cuantitativamente el número de conductas acaecidas en las clases.

Posteriormente, apareció otro sistema de categorías para observar las conductas antisociales (CAS), diseñada por Gras y Martínez (2002), que evaluaba la conducta de los escolares mientras realizaban actividad física. Este instrumento agrupa las conductas antisociales en seis categorías: agresión, quitar (robar), molestar, falta de control, forzar u obligar y la categoría otros. Se trata de un instrumento de observación con ciertas limitaciones, como la exclusión en la categoría de “agresión” de las vertientes verbales y una definición poco operativa de la categoría “otros”. Por otro lado, este sistema de categorías presenta un diseño específico para la observación “in situ” y, tal y como señala Anguera (2010), las tendencias actuales en investigación observacional tratan de evitar este tipo de observaciones para avanzar hacia la grabación de las conductas y su análisis informatizado, con el objeto de poder volver a observar una conducta que ya ha sucedido, o poder registrar varias conductas que acaecen simultáneamente.

En este sentido, Escartí et al. (2005) elaboraron un sistema de categorías para la evaluación de los comportamientos de responsabilidad personal y social en el que, a través de la visualización del video de las clases de Educación Física, los observadores identifican de manera cuantitativa la aparición de las conductas de agredir, interrumpir, no colaborar y ayudar a lo largo de diferentes sesiones. Este sistema no ha sido validado científicamente: presenta una concordancia consensuada y no profundiza en los análisis de calidad de dato (Hernández, Díaz y Morales, 2010).

Estos mismos autores expusieron una nueva herramienta observacional que contenía altos valores de fiabilidad y validez, específico para las clases de Educación Física, pero registraba 22 conductas que alteran la convivencia, un número muy elevado que dificultaba su aplicación, además de que no incluía conductas tan importantes como las agresiones.

Por lo tanto, el limitado número de instrumentos de observación de conductas que alteran la convivencia en las clases de Educación Física, y los problemas

existentes en los ya creados, tales como un elevado o bajo número de conductas a observar, registros cualitativos, técnicas de observación *in situ* o la no validación científica de alguna de ellas, ponen de manifiesto la necesidad del diseño de un nuevo instrumento que supla las carencias de las ya existentes.

De este modo, el objetivo principal de esta investigación fue diseñar y validar un instrumento de observación de conductas que alteran la convivencia en las clases de Educación Física que permita definirlas y categorizarlas, facilitando su aplicación sin la presencia del observador en el aula y minimizando el efecto de esa presencia en el comportamiento del alumnado.

MÉTODO

Muestra

Para el desarrollo de la investigación se emplearon dos tipos de participantes. En primer lugar, se solicitó la colaboración de seis jueces expertos, todos ellos Maestros especialistas en Educación Física, Licenciados en Ciencias de la Actividad Física y el Deporte, docentes en activo y Doctores en Ciencias de la Actividad Física y el Deporte o Catedráticos de Educación Secundaria.

En segundo lugar, cuatro observadores, Licenciados en Ciencias de la Actividad Física y el Deporte, fueron formados en la valoración de las conductas que alteran el clima de convivencia, y se encargaron de visualizar el número de conductas y debatir sobre la tipología de las mismas en el proceso de validación del instrumento.

Para ello, se utilizaron un total de 10 clases de Educación Física de cursos de sexto de Educación Primaria y tercero de Educación Secundaria Obligatoria, de 10 Centros Escolares diferentes. En este sentido, se analizó la frecuencia de las conductas que alteran la convivencia de 10 sesiones de aproximadamente 50 minutos de duración, que fueron utilizadas para el cálculo de las pruebas de fiabilidad intra-observador, inter-observador, resultados descriptivos, consistencia interna y análisis factorial exploratorio y confirmatorio. La selección del muestreo de los centros fue por conveniencia, debido a la accesibilidad por parte de los investigadores. Los centros se encontraban en diferentes contextos socio-económicos. Las clases analizadas estaban compuestas por un número de entre 20 y 32 alumnos. En este caso, la selección de las clases dentro de cada centro educativo fue aleatoria dentro de los cursos elegidos en el estudio (sexto de Educación Primaria y tercero de Educación Secundaria).

Procedimiento

El procedimiento de la investigación estuvo formado por varias fases, en las que se realizaron diferentes versiones del instrumento definitivo. Durante este proceso, se siguió la línea de trabajo de los autores Cruz et al. (1996) en la validación de un instrumento de observación de conductas relacionadas con el *fair play* en fútbol.

Fase 1. Elaboración de la primera versión del Instrumento de Observación de Conductas que Alteran la Convivencia en las clases de Educación Física (CACEF)

En esta primera fase se elaboró una versión a partir del instrumento de Díaz y Esteban (2010), formado por 59 ítems que describen conductas inadecuadas, clasificadas en seis categorías, relacionadas con la agresividad, la falta de madurez, el desinterés académico, las conductas disruptivas, antisociales y de indisciplina. Dicho instrumento es una adaptación transcultural del *Physical Education Classroom Management Instrument* (Instrumento de la Gestión del Aula en Educación Física), desarrollado por Kulinna, Cothran y Recualos (2003).

Posteriormente, un grupo formado por seis jueces expertos buscó reducir el número de ítems de forma que fueran los más significativos, por un lado, y que se pudieran utilizar de una forma eficaz durante las sesiones de Educación Física, por otro, facilitando así el entrenamiento de los observadores. El resultado fue un instrumento de observación inicial con 12 ítems encaminados a determinar la frecuencia con la que se producían las conductas por parte de cualquiera de los alumnos en el desarrollo de la sesión (Tabla 1).

Tabla 1. Conductas utilizadas para confeccionar la primera versión del CACEF

CONDUCTAS

1. Quejarse	7. No seguir las indicaciones del profesor
2. Hacer trampas, engañar en los juegos y deportes	8. Interrumpir o molestar a los demás
3. Reírse, burlarse de los compañeros	9. Insultar, agredir verbalmente
4. Agredir físicamente	10. Realizar gestos obscenos
5. Excluir a compañeros de juegos y deportes	11. Amenazar a un compañero
6. Recriminar, discutir irrespetuosamente	12. Maltratar el material

Fuente: elaboración propia.

Fase 2. Ensayo, discusión y reestructuración de la primera versión del CACEF

En esta segunda fase, cuatro observadores analizaron en común dos sesiones de 60 minutos de una clase de Educación Física de sexto curso de Educación Primaria y de tercero de Educación Secundaria Obligatoria. De este modo, se realizó una primera prueba del instrumento de observación en la que se analizaron y registraron las conductas, después de una puesta en común por parte de los observadores.

Cuando se finalizó con el análisis del video, y tras comentar los aspectos más relevantes sobre la facilidad o dificultad de registrar aquello observado en el monitor, se realizó una puesta en común para añadir, eliminar, modificar o agrupar las conductas de la plantilla inicial, elaborando una segunda versión.

- Conductas agrupadas: fueron la conducta “agredir verbalmente”, la conducta amenazar y la conducta “reírse y burlarse de los compañeros” agrupadas en una única conducta, denominada Agredir Verbalmente, que ocupa el primer lugar en el instrumento de registro.

- Conductas renombradas: se renombró la conducta “realizar gestos obscenos”, que pasó a llamarse “agredir gestualmente” y ocupó el tercer lugar en el instrumento de registro.

- Conductas eliminadas: se eliminó la conducta “recriminar/discutir irrespetuosamente”, por considerar que podría suponer motivo de confusión con la conducta “agredir verbalmente” y la conducta “quejarse”.

Fase 3. Ensayo, discusión y realización del instrumento definitivo

En esta ocasión, se utilizó la segunda versión de la plantilla para el análisis de otras dos sesiones de 50 minutos de otras dos clases de Educación Física. Posteriormente, en una sesión de laboratorio, se pusieron en común los registros obtenidos por los cuatro observadores con el fin de conocer las coincidencias y discrepancias entre ellos, así como la objetividad del instrumento de observación. Se determinó que, únicamente en el caso de confusión y/o discrepancia entre los cuatro observadores, se volvía a repetir la acción hasta llegar a un acuerdo sobre cuál debía ser la conducta o categoría a la que la acción debía ser adscrita. Tras la discusión sobre los registros obtenidos, y la utilidad práctica de la plantilla, se decidió realizar algunas modificaciones en ésta:

- Conductas renombradas: se renombró la conducta “maltratar el material”, que pasó a denominarse “maltratar el material o las instalaciones” y no abarcaba exclusivamente el material de clase, sino también las instalaciones del centro.

En esta fase, una vez obtenido el modelo del instrumento final, se llevó a cabo la definición operacional de todas las conductas del instrumento de observación (Tabla 2), redefiniendo aquellas que habían planteado problemas a los observadores.

Tabla 2. Definición de conductas del instrumento de observación

CONDUCTAS	DEFINICIÓN
Agredir físicamente	Acciones del alumnado en las que, más allá de las situaciones deportivas en las que pueda existir contacto físico, los alumnos empujan, dan patadas, puñetazos, golpes, zancadillean o zarandean a otros compañeros con una clara mala intención en la acción. Ejemplos: “un alumno golpea a otro”, “estando agrupados en hilera un compañero empuja al de delante para que contacte a su vez con el siguiente”.
Agredir verbalmente	Acciones del alumnado en las que se mofan de los compañeros o del profesor, los insultan, amenazan o faltan al respeto mediante la palabra. Ejemplos: “cállate, eres tonto”, “deja de molestarme, eres un patoso”.
Agredir gestualmente	Situaciones en las que los alumnos se expresan con gestos inapropiados, ofensivos o con connotaciones sexuales. Por ejemplo: “un alumno realiza un corte de mangas”. También puede considerarse agresión gestual aquella conducta por la cual el alumno amenaza con pegar o maltratar a un compañero, aun sin llegar a hacerlo. Por ejemplo: “un alumno alza el puño para pegar a otro, pero sin llegar al contacto”.
Interrumpir o molestar a los demás	Acciones en las que los alumnos no dejan que la clase tenga su continuidad normal, ya sea durante la realización de las actividades o durante la explicación de las mismas. Se registrará como conducta de interrumpir o molestar a los demás, cada vez que el profesor interrumpa la clase pidiendo silencio o llamando la atención a algún alumno. Si el profesor pide silencio de forma general a toda la clase, se anotará como una única conducta de interrupción, pero si lo hace de forma específica, por ejemplo, a tres alumnos en concreto, se registrará como tres conductas de interrupción. Ejemplos: “un alumno comienza a hablar o a molestar a sus compañeros en el momento en el que el profesor explica una tarea”, “los alumnos agrupados en parejas están pasándose un balón y un tercero le sustrae el balón a una de las parejas”.
Maltratar el material o a las instalaciones	Se producen en los momentos en los que el alumno utiliza de manera irresponsable las instalaciones o el material, con excesiva fuerza, o realizando tareas para las que no están diseñados y que pueden provocar que se deterioren. Ejemplos: “el alumno da una patada a los conos de demarcación”, “escupir en el suelo”.
Hacer trampas, engañar en los juegos y deportes	Se produce cuando los alumnos incumplen las normas de juegos y actividades con ánimo de conseguir la victoria o alguna ventaja para determinadas acciones, o durante el desarrollo de un juego o actividad, o para dar ventaja a su equipo con estas acciones. Ejemplos: “los alumnos cuentan más goles en el partido de los que se han producido”, “los jugadores acaban la actividad sin que todos la hayan llevado a cabo, norma que había establecido el profesor para llegar a su fin”.

[CONTINÚA PÁGINA SIGUIENTE]

Excluir a compañeros	Se produce en las situaciones en las que el alumnado no deja participar, ignora o minusvalora a algunos compañeros a la hora de realizar tareas propuestas por el profesor en las que se requiere la formación de equipos, grupos o parejas. Ejemplos: “un alumno queda sin equipo cuando se proponen para realizar una tarea”, “no pasan la pelota a un compañero debido a que su nivel de habilidad es menor que el del resto del equipo y puede suponerle una desventaja a la hora de obtener mejor puntuación”.
Quejarse	Se produce cuando el alumno expresa su descontento con la actividad o con los compañeros. Debe cumplir dos requisitos para que se considere una conducta que altera la convivencia: que no exista motivo objetivo para la queja y que se produzca de manera reiterada por el mismo alumno, expresando su reclamación con salidas de tono, sin llegar a faltar al respeto, insultar o amenazar, en cuyo caso sería una conducta de agresión verbal. Ejemplos: “estoy cansado de hacer lo mismo siempre”, “¡qué aburrimiento!, ¿cuándo vamos a poder jugar a lo que nosotros queremos?”.
No seguir las indicaciones del profesor	Situaciones en las que el alumnado ignora, realiza de otra manera, o no realiza los ejercicios y tareas propuestos por el profesor, o no sigue las consignas propuestas para la sesión o tarea en concreto. Ejemplos: “un alumno comienza a tirar a puerta, cuando el profesor ha indicado que hay que realizar pases con el compañero”, “el profesor ha solicitado que se recoja el material y el alumno sigue jugando”. Por otro lado, se registrarán como “no seguir las indicaciones” tantas conductas realizadas por el alumno, pero si esta conducta es de forma continuada en un mismo ejercicio, sólo se anotará una vez. Por ejemplo: “un alumno no participa en la actividad de fútbol” constará como una única conducta. Para considerarse conducta disruptiva, debe pasar un tiempo mínimo de 10 segundos seguidos en los que el alumno no realice la actividad indicada por el profesor, ya que, en ocasiones, algunos alumnos pueden pararse debido a que están cansados o están cambiando de un ejercicio a otro diferente. Es importante anotar esta conducta únicamente cuando el comportamiento del alumno no se corresponda a ninguna de las conductas anteriores. En este ejemplo: “un alumno empuja a otro durante un juego”, aunque el alumno no sigue la indicación del profesor de no empujar violentamente, la conducta que se ha de registrar es una agresión física.

Fuente: elaboración propia.

Finalmente, se añadieron apartados que mejoraban y ampliaban la información facilitada por el instrumento CACEF, tales como secciones para conocer la fecha de observación de las clases de Educación Física, anotar el nombre del centro, el curso, el número de alumnos de la clase, el tiempo total de la sesión y una sección de observaciones.

Fase 4. Juicio de expertos y análisis de los datos

El objetivo principal de esta fase fue adquirir la validez de las diferentes conductas que formaban el CACEF. Para llevar a cabo esta validación, el instrumento fue enviado a un panel de seis jueces expertos que adjuntaba una carta en la que se

justificaba su colaboración y el conocimiento del objeto de estudio. Dichos jueces observaron el grado de representatividad y calidad técnica de todas las conductas que forman el instrumento, realizando importantes aportaciones y valoraciones de acuerdo con la metodología de Ortega, Jiménez, Palao y Sainz de Baranda (2008). Para considerar si las conductas son adecuadas para realizar las mediciones, los jueces expertos realizaron una revisión cualitativa del instrumento de observación mediante el Método Delphi (Landeta, 1999), del mismo modo que evaluaron cuantitativamente los diferentes aspectos de los ítems mediante una escala Likert de 10 categorías. Se solicitó a los jueces expertos que valorasen los diferentes aspectos relativos a: a) representatividad, o grado en que el ítem tal y como está planteado es el mejor de los posibles; y b) calidad técnica o grado en que el ítem, tal y como está redactado, no induce a error por sesgos gramaticales. Posteriormente, se calculó la fiabilidad, definida como reproducibilidad de una medida (Thomas, Silverman y Nelson, 2015). De este modo, se calculó la fiabilidad interobservadores y la fiabilidad intraobservador mediante la prueba estadística Coeficiente de Correlación Intraclase (CCI) y la prueba test-retest. Además, se calcularon los estadísticos descriptivos y la consistencia interna (Media, desviación típica, coeficiente de correlación corregido ítem-total, alfa de Cronbach e índices de asimetría y curtosis). Finalmente, con objeto de agrupar las ocho conductas restantes en diferentes categorías o dimensiones, se realizó un análisis factorial exploratorio con máxima verosimilitud; posteriormente se aplicaron modelos de ecuaciones estructurales, evaluando la estructura factorial del instrumento mediante AFC. Se utilizó el programa Amos 18.0 y se obtuvieron los parámetros a través del método de máxima verosimilitud. Para evaluar el ajuste de los modelos confirmatorios se utilizaron varios índices, absolutos y relativos, como recomiendan Bentler (2007) y Markland (2007). De este modo, para los índices absolutos se utilizó el valor p , asociado con el estadístico chi cuadrado (X^2), así como la ratio entre el índice de X^2 y los grados de libertad (gl), considerando los valores por debajo de 2,0 como indicadores de muy buen ajuste de modelo (Tabachnik y Fidell, 2007). Asimismo, se calculó el GFI (índice de bondad de ajuste), y, entre los índices relativos, el NFI (*índice de ajuste normalizado*), el NNFI (*índice de ajuste no normativo*) y el CFI (*índice de ajuste comparativo*), cuyos valores deben ser iguales o superiores a 0,90 para considerar aceptable el ajuste de un modelo (Kaplan, 2000). Autores como Kline (2005) recomiendan la utilización del RMSEA (error de aproximación cuadrático medio) y, según Steiger (2007), un valor menor a 0,07 indicaría un buen ajuste.

RESULTADOS

Validez del instrumento CACEF mediante jueces expertos

En primer lugar, se realizó la media de las valoraciones cuantitativas con respecto a la representatividad y calidad técnica de los ítems; no se adoptó ninguna medida adicional porque todos los ítems obtuvieron puntuaciones medias superiores a seis (Osterlind, 1989).

Con respecto a las valoraciones cualitativas, todos los jueces expertos coinciden afirmando que el instrumento CACEF cumple con el grado de comprensión, adecuación y definición requeridas para su uso, aunque consideran necesario un exhaustivo entrenamiento por parte de los observadores que les permita registrar las mismas conductas al analizar los videos, debido a la subjetividad de las conductas.

Por otro lado, en la Tabla 3 se puede observar que la valoración del instrumento de observación se completa con la prueba V de Aiken (Merino y Livia, 2009).

Tabla 3. Valoración cuantitativa de los expertos mediante la prueba V de Aiken

ÍTEMS - CONDUCTAS	V DE AIKEN	
	REPRESENTATIVIDAD	CALIDAD TÉCNICA
1. Agredir verbalmente	.92	.77
2. Agredir físicamente	.97	.94
3. Agredir gestualmente	.91	.77
4. Interrumpir o molestar a los demás	.94	.91
5. Maltratar el material o las instalaciones	.88	.88
6. Hacer trampas, engañar en los juegos y tareas	.88	.86
7. Excluir a compañeros	.91	.77
8. Quejarse	.91	.88
9. No seguir las indicaciones del profesor	.88	.80
Global	.91	.84

Fuente: elaboración propia.

Fiabilidad

Evaluación de la fiabilidad interobservadores

Para evaluar la concordancia entre observadores, se contó con la colaboración de tres nuevos observadores formados en la observación de conductas que alteraban

el clima de aula en Educación Física. En este proceso se explicó detenidamente el instrumento y se analizaron 10 sesiones, tras las cuales se hacía una puesta en común de los acuerdos en el número y tipos de conductas observadas. En aquellas conductas en las que no existía acuerdo, se volvía a analizar la escena y se exponían las causas por las que unos observadores sí habían detectado alguna conducta y por qué otros no, elaborando un protocolo para sucesivas observaciones. Tras un número aproximado de cinco sesiones de entrenamiento se procedió a analizar el grado de acuerdo entre los observadores, para lo cual se empleó la prueba estadística denominada Coeficiente de Correlación Intraclase, obteniendo un valor global de 0,89, considerado como muy aceptable.

- Evaluación de la fiabilidad intraobservador

Para el cálculo de la prueba test-retest se utilizaron dos de las sesiones de Educación Física de 45 minutos de duración, cada una de ellas observadas en el proceso de evaluación de la fiabilidad inter-observador: una con alumnos de sexto de Educación Primaria y otra con alumnos de tercero de la ESO. De este modo, tres nuevos investigadores, previamente formados en la valoración de las conductas que alteran el clima del aula en las clases de Educación Física, visualizaron ambas sesiones en el mismo día, anotando las diferentes conductas que fueron observando, y volvieron a repetir este proceso tras un intervalo de una semana. Finalmente, se realizó el cálculo del coeficiente de correlación intraclase: la cifra media fue de .98, lo que indica un grado de concordancia adecuado.

Estadísticos descriptivos y consistencia interna

Las características de los ítems que forman el CACEF fueron analizadas en el total de las 10 sesiones de Educación Física analizadas (Tabla 4) y se comprobó si el alfa de la escala aumentaba con la eliminación de algún ítem. El coeficiente de correlación corregido ítem-total (CCIT-c) ³ -.30 y desviación típica (DT) > 1. Por lo tanto, el ítem siete, “excluir a compañeros”, fue eliminado del instrumento de observación, debido a los bajos valores presentados. Como se puede comprobar, y según las recomendaciones de Bollen y Long (1994), los índices de asimetría y curtosis son próximos a 0 y < 2 en la mayoría de los ítems, exceptuando el ítem ocho, “quejarse”, que presentó unos valores elevados, y el ítem siete, “excluir a compañeros”, que fue eliminado. Por lo tanto, la escala final estuvo formada por ocho conductas, mostrando un Alfa de Cronbach de .73, valor aceptable para este tipo de investigaciones.

Tabla 4. Estadísticos descriptivos y consistencia interna

ESCALA:	M	DT	CCIT-C	A SIN ÍTEM	ASIMETRÍA	CURTOSIS
1. Agredir verbalmente	3.07	2.14	.66	.68	1.19	1.54
2. Agredir físicamente	3.14	2.27	.50	.69	1.21	2.13
3. Agredir gestualmente	.79	.99	.51	.71	1.68	2.25
4. Interrumpir o molestar a los demás	14.25	8.09	.75	.62	.08	-.97
5. Maltratar el material o las instalaciones	.68	.90	.49	.71	1.35	1.29
6. Hacer trampas, engañar en los juegos	13.14	4.58	.41	.69	.29	-.51
7. Excluir a compañeros	.36	.73	.05	.73	2.37	2.91
8. Quejarse	1.93	1.36	.34	.71	.58	-.635
9. No seguir las indicaciones del profesor	14.21	7.78	.72	.62	.15	-.303

Fuente: Elaboración propia.

Análisis Factorial Exploratorio

Se realizó el cálculo del análisis factorial exploratorio teniendo en cuenta lo expuesto anteriormente en este trabajo en relación a la consistencia interna (Tabla 4) y eliminando el ítem siete, “excluir a los compañeros”. Entre las salidas obtenidas se encuentra el índice KMO (Kaiser-Meyer-Olkin), que compara los coeficientes de correlación lineal con los coeficientes de correlación parcial. Si los coeficientes de correlación parcial entre las variables son muy pequeños, quiere decir que la relación entre cada par de las mismas se debe o puede ser explicada por el resto, y por tanto se puede llevar a cabo un análisis factorial.

Calculada la medida KMO, el resultado obtenido fue de .65, que está dentro del límite a partir del cual el análisis factorial a realizar es adecuado y las conclusiones consistentes. Al desarrollar la prueba de esfericidad de Bartlett se obtiene un p-valor < .001, lo que hace rechazar la hipótesis de que la matriz de correlaciones es una matriz de identidad, confirmándose una de las hipótesis de partida. El determinante de la matriz de correlaciones ofrece un valor muy bajo y muchas de estas correlaciones son significativas, apoyando la idea de que el análisis factorial es adecuado.

Al realizar el análisis factorial exploratorio se obtienen dos factores o dimensiones que explican el 68.03% de la variabilidad total, lo que supone un porcentaje bastante aceptable, obteniendo los grupos de ítems que saturan cada factor una vez rotados. Así, se obtiene un primer factor formado por los ítems “agredir gestual-

mente”, “agredir verbalmente”, “agredir físicamente” y “maltratar el material o las instalaciones”, que hacen referencia a la categoría de conductas violentas. El factor dos, “no seguir las indicaciones del profesor”, “quejarse”, “hacer trampas, engañar en los juegos y deportes” e “interrumpir o molestar a los demás”, hace referencia a la categoría de indisciplina o antisociales. De este modo, las conductas se agrupan en dos factores o categorías: conductas violentas y conductas de indisciplina o desinterés (Tabla 5), siguiendo la línea de los autores Kulinna et al. (2003).

Tabla 5. Factores obtenidos y saturación de los ítems en el factor

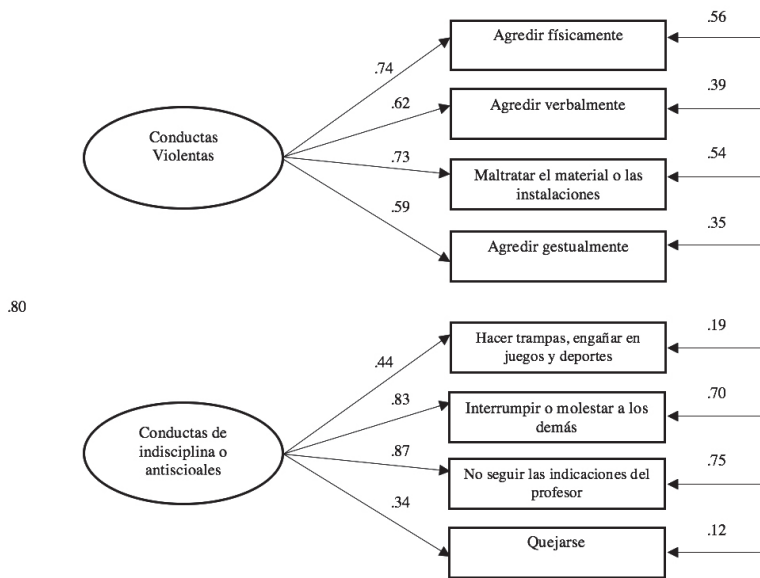
CONDUCTAS VIOLENTAS	SATURACIÓN	CONDUCTAS DE INDISCIPLINA O DESINTERÉS	SATURACIÓN
Agredir verbalmente	.52	Interrumpir o molestar a los demás	.94
Agredir físicamente	.59	Hacer trampas, engañar en los juegos	.37
Agredir gestualmente	.99	Quejarse	.27
Maltratar el material o las instalaciones	.50	No seguir las indicaciones del profesor	.60

Fuente: elaboración propia.

Análisis Factorial Confirmatorio

A continuación, en la Figura 1, se presentan los resultados del AFC correspondiente al modelo de ocho conductas que alteran la convivencia en las clases de Educación Física. Los índices de bondad mostraron un ajuste satisfactorio: $X^2 = 26.78$, $gl = 19$, $X^2/gl = 1.40$, $GFI = .84$, $CFI = .88$, $RMSEA = .12$.

Figura 1. Diagrama de la trayectoria (Path Diagram) del AFC, con pesos estandarizados y errores de medición de cada una de las conductas del CACEF



DISCUSIÓN

El objetivo del presente estudio ha sido diseñar y validar un instrumento de observación para el análisis de las conductas que alteran la convivencia en las clases de Educación Física. Los resultados obtenidos en la investigación muestran que el diseño y elaboración de un instrumento de observación requiere de diversas fases, así como colaboración de jueces expertos (Carretero-Dios y Pérez, 2005; Cruz et al., 2006; Kulinna et al., 2003).

El número de jueces expertos consultado fue superior al empleado en otros trabajos similares (Baena, Granero, Brancho y Pérez, 2012) y realiza así importantes aportaciones para la mejora del instrumento de observación, que giran en torno a la definición, agrupación y objetividad de las conductas observadas. Dichas contribuciones de tipo cualitativo son indispensables en el desarrollo de un instrumento (Carretero-Dios y Pérez, 2005; Ortega et al., 2008). Por otro lado, los valores cuantitativos otorgados por el panel de expertos fueron muy elevados y se reflejan en la prueba V de Aiken, donde se aprecia que los resultados obtenidos para cada uno de los ítems, tanto en representatividad como calidad técnica, son muy eleva-

dos, superiores a los mínimos propuestos por Penfield y Giacobbi (2004), tanto en la escala global como en cada una de las conductas.

Los resultados de la consistencia interna del instrumento otorgaron un coeficiente alfa de Cronbach aceptable para este tipo de investigaciones (Morales, 2007), y similar a la obtenida por otros autores en este tipo de instrumentos (Baena et al., 2012; Cruz et al., 2006; Díaz y Esteban, 2010). Además, los resultados referentes a la fiabilidad intra-observador e inter-observador obtuvieron valores considerados como muy aceptables, y en concordancia a los obtenidos en otras investigaciones similares (Cruz et al., 2006; Gálvez-Nieto, Vera-Bachman, Cerda y Díaz, 2016).

Por último, los análisis psicométricos y su correspondiente análisis factorial confirmatorio basado en ecuaciones estructurales apoyan la validez factorial y la fiabilidad del instrumento, formado por ocho conductas que alteran la convivencia en las clases de Educación Física, agrupadas en dos dimensiones (conductas violentas y conductas de indisciplina o desinterés). De este modo, el número de conductas a evaluar y de ítems que contiene este instrumento se halla en un término medio con respecto al resto de instrumentos desarrollados hasta el momento. Contiene un número inferior a los instrumentos creados por Cruz et al. (2006), Esteban y Díaz (2010), Auné, Abal y Attorresi (2016), formado por 59 conductas, pero mayor en comparación con los empleados en los estudios de Escartí et al. (2005) y de Gras y Martínez (2002).

De este modo, el instrumento de observación de conductas que alteran la convivencia en las clases de Educación Física ha obtenido altos valores en los índices de fiabilidad (inter e intraobservador y consistencia interna) y validez, para aquellas clases de Educación Física filmadas en vídeo, así como en el análisis factorial exploratorio y confirmatorio. Por lo tanto, este instrumento permitirá analizar de forma cuantitativa, a través de la observación indirecta, el número y tipología de conductas que se producen en las aulas, que permitirá al docente conocer el nivel de convivencia en sus clases y comparar de manera objetiva con las aulas de otros centros. Sin embargo, este estudio presenta ciertas limitaciones que deben ser tomadas en cuenta a la hora de interpretar los resultados. En ese sentido, este instrumento de observación no garantiza la exhaustividad y mutua exclusividad de las categorías, ya que las ocho categorías del CACEF no son mutuamente excluyentes (por ejemplo, es posible agredir verbalmente a la vez que físicamente). Igualmente, hubiese sido interesante haber involucrado al profesorado activo, que posteriormente podía utilizar el instrumento además de los observadores formados y especializados.

Finalmente, se puede concluir que, tras este procedimiento encaminado a la creación y validación de un instrumento para registrar las conductas que alteran la convivencia en las clases de Educación Física, los resultados son satisfactorios.

Se trata de un sistema de categorías formado por ocho ítems agrupados en dos dimensiones que se consideran exhaustivas y mutuamente excluyentes, y que permite recoger un amplio abanico de conductas que alteran la convivencia y abarca, desde una visión integradora, la tipología de comportamientos que hace que el ambiente de clase se vea perjudicado.

Fecha de recepción del original: 24 de septiembre 2017

Fecha de aceptación de la versión definitiva: 26 de abril 2018

REFERENCIAS

- Anguera, M. T. (2010). Posibilidades y relevancia de la observación sistemática por el profesional de la psicología. *Papeles del Psicólogo*, 31(1), 122-130.
- Anguera, M. T., Blanco, A., Hernández, A. y Losada, J. L. (2011). Diseños observacionales: ajuste y aplicación en psicología del deporte. *Cuadernos de Psicología del Deporte*, 11(2), 63-76.
- Auné, S. E., Abal, F. J. y Attorresi, H. F. (2016). Diseño y construcción de una Escala de Conducta Prosocial para adultos. *Revista Iberoamericana de Diagnóstico y Evaluación Psicológica*, 42(2), 15-25.
- Baena, A., Granero, A., Brancho, C. y Pérez, F. J. (2012). Versión española del Sport Satisfaction Instrument (SSI) adaptado a la Educación Física. *Revista de Psicodidáctica*, 17(2), 377-396.
- Barnes, T. N., Smith, S. W. y Miller, M. D. (2014). School-based cognitive-behavioral interventions in the treatment of aggression in the United States: A meta-analysis. *Aggression and Violent Behavior*, 19, 311-321.
- Benítez, J. L., Tomás, A. y Justicia, F. (2005). Educación para la convivencia en contextos escolares: una propuesta de intervención contra los malos tratos entre iguales. *Apuntes de Psicología*, 23(1), 27-40.
- Bentler, P. M. (2007). On tests and indices for evaluating structural models. *Personality and Individual Differences*, 42, 825-829.
- Bollen, K. A. y Long, J. (1994). *Testing structural equation models*. Newbury Park: Sage.
- Borràs, A., Palou, P. y Ponseti, F. X. (2001). Evaluación del fairplay y conductas violentas durante la práctica deportiva de estudiantes de magisterio en Educación Física. En J. F. Campos, S. Llanas y R. Aranda (Coords.), *Nuevas aportaciones al estudio de la actividad física y el deporte*. II Congreso de CAFyD. Valencia: Marí Montañana.
- Carretero-Dios, H. y Pérez, C. (2005). Normas para el desarrollo y revisión de

- estudios instrumentales. *International Journal of Clinical and Health Psychology*, 5(3), 521-551.
- Cruz, J., Capdevila, L., Boixados, M., Pintanel, M., Alonso, C., Mimbbrero, J. y Torregrosa, M. (1996). Identificación de conductas, actitudes y valores relacionados con el fairplay en deportistas jóvenes. En *Valores sociales y deporte. Fair Play versus violencia. Investigaciones en ciencias del deporte* (pp. 37-67). Madrid: Consejo Superior de Deportes.
- Díaz, A. y Esteban, R. (2010). Physical Education Classroom Management Instrument. Adaptación transcultural. *Cuadernos de Psicología del Deporte*, 10, 73-78.
- Escartí, A., Pascual, C. y Gutiérrez, M. (2005). *Responsabilidad personal y social a través de la educación física y el deporte*. Barcelona: Graó.
- Fernández, A. B. y Barreiras, A. J. (2015). Stop-Motion: Estrategia de Imagen y Comunicación para la mejora de la convivencia. *Revista de Estudios e Investigación en Psicología y Educación*, 13, 191-195.
- Gálvez-Nieto, D., Vera-Bachman, D., Cerda, D. y Díaz, R. (2016). Escala de victimización entre adolescentes a través del teléfono móvil y de internet: Estudio de validación de una versión abreviada en estudiantes chilenos. *Revista Iberoamericana de Diagnóstico y Evaluación Psicológica*, 41(1), 16-27.
- García, L. (2010). Los planes de convivencia de centro. Reflexiones y orientaciones. En J. J. Gázquez y M. C. Pérez (Eds.), *La convivencia escolar. Aspectos psicológicos y educativos* (pp. 73-79). Granada: GEU.
- Gras, M. y Martínez, G. (2002). Las primeras manifestaciones de la conducta antisocial en la escuela. *Revista Electrónica Interuniversitaria de Formación del Profesorado*, 5(4), 15.
- Goyette, R., Dore, R. y Dion, E. (2000). Pupils' misbehaviour and the reactions and causal attributions of physical education student teachers: A sequential analysis. *Journal of Teaching in Physical Education*, 20, 3-14.
- Hemphill, M. A., Templin, T. J. y Wright, P. M. (2015). Implementation and outcomes of a responsibility-based continuing professional development protocol in physical education. *Sport, Education and Society*, 20(3), 398-419.
- Hernández, 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.
- Kaplan, D. (2000). *Structural equation modeling: Foundations and extensions*. Thousand Oaks: Sage.
- Kuhn, T. M., Ebert, J. S., Gracey, K. A., Chapman, G. L. y Epstein, R. A. (2015). Evidence-based interventions for adolescents with disruptive behaviors in

- school-based settings. *Child and Adolescent Psychiatric Clinics of North America*, 24(2), 305-317.
- Kulinna, P. H., Cothran, D. y Recualos, R. (2003). Development of an instrument to measure disruptive behaviour. *Measurement in Physical Education and Exercise Science*, 7(1), 25-41.
- Kline, R. B. (2005). *Principles and practice of structural equation modeling*. New York: The Guilford Press.
- Landeta, J. (1999). *El método Delphi. Una técnica de previsión para la incertidumbre*. Barcelona: Ariel.
- Markland, D. (2007). The golden rule is that there are not golden rules: A commentary on Paul Barrett's recommendations for reporting model fit in structural equation modelling. *Personality and Individual Differences*, 42, 851-858.
- Merino, C. y Livia, S. (2009). Intervalos de confianza asimétricos para el índice de validez de contenido: un programa Visual Basic para la V de Aiken. *Anales de Psicología*, 25(1), 169-171.
- Morales, P. (2007). *El contraste de medias. Diferencia entre medias de muestras relacionadas*. Madrid: Universidad Pontificia de Comillas.
- Ortega, E., Jiménez, J. M., Palao, M. y Sainz de Barranda, P. (2008). Diseño y validación de un cuestionario para valorar las preferencias y satisfacciones en jóvenes jugadores de baloncesto. *Cuadernos de Psicología del Deporte*, 8(2), 39-58.
- Osterlind, S. J. (1989). *Constructing test items*. Boston: Kluwer.
- Palou, P., Borràs, P. A., Ponseti, F. X., Vidal, J. y Torregrosa, M. (2007). Intervención para la promoción de actitudes de fairplay en futbolistas cadetes. *Apunts*, 89, 15-22.
- Penfield, R. D. y Giacobbi, P. R. (2004). Applying a score confidence interval to Aiken's item content-relevance index. *Measurement in Physical Education and Exercise Science*, 8(4), 213-225.
- Peñalva, A., López-Goñi, J. J. y Landa, N. (2013). Competencias emocionales del alumnado de Magisterio: posibles implicaciones profesionales. *Revista de Educación*, 362, 690-712.
- Ruiz, L. M., Rodríguez, P., Martinek, T., Schilling, T., Durán, L. J. y Jiménez, P. (2006). El proyecto esfuerzo: un modelo para el desarrollo de la responsabilidad personal y social a través del deporte. *Revista de Educación*, 341, 933-958.
- Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modelling. *Personality and Individual Differences*, 42, 893-898.
- Tabachnick, B. G. y Fidell, L. S. (2007). *Using multivariate statistics*. New York: Allyn and Bacon.

- Thomas, J. R., Silverman, S. y Nelson, J. (2015). *Research methods in physical activity*, Champaign: Human Kinetics.
- Urbina, C., Simón, C. y Echeita, G. (2011). Concepciones de los profesores acerca de las conductas disruptivas: análisis a partir de un marco inclusivo. *Infancia y Aprendizaje*, 34(2), 205-217.
- Whitley, M. A., Forneris, T. y Barker, B. (2015). The reality of sustaining community-based sport and physical activity programs to enhance the development of underserved youth: Challenges and potential strategies. *Quest*, 67(4), 409-423.

ANEXO 1

VERSIÓN DEFINITIVA DEL INSTRUMENTO DE OBSERVACIÓN DE CONDUCTAS QUE ALTERAN LA CONVIVENCIA EN CLASE DE EDUCACIÓN FÍSICA

CENTRO: _____
 FECHA: ____/____/____
 GRUPO: _____ N° ALUMNOS: _____ TIEMPO: _____ min _____ seg _____

CONDUCTA	MOMENTO DE APARICIÓN DE LA CONDUCTA (Tiempo: min. y seg.)												FRECUENCIA TOTAL	
1. Agredir verbalmente														
2. Agredir físicamente														
3. Agredir gestualmente														
4. Interrumpir o molestar a los demás														
5. Maltratar el material o las instalaciones														
6. Hacer trampas, engañar en los juegos y deportes														
7. Quejarse														
8. No seguir las indicaciones del profesor														

OBSERVACIONES: