

Construcción y validación del Cuestionario de Acoso y Ciberacoso Escolar (CACE) mediante juicio de expertos

Construction and Validation of the School Bullying and Cyberbullying Questionnaire (CACE) through Expert Judgment

Cristián David Expósito *¹, Roxana Graciela Marsollier²,
Hilda Difabio de Anglat¹, Alejandro Castro-Santander³

1 - Centro de Investigaciones Cuyo – CONICET, Mendoza, Argentina.

2 - CONICET – Universidad Nacional de Cuyo, Mendoza, Argentina.

3 - Observatorio de la Convivencia Escolar – Universidad Católica de Cuyo, Mendoza, Argentina.

Desarrollo
Metodología
Resultados
Discusión
Referencias
Anexo

Recibido: 06/02/2023 Revisado: 10/02/2023 Aceptado: 13/02/2023

Resumen

El objetivo del presente artículo es validar los ítems de una prueba psicométrica sobre acoso y ciberacoso a través de la constitución de un tribunal de expertos. Se utilizó una metodología mixta con predominancia cuantitativa basada en la demostración de cuatro criterios teóricos, es decir, suficiencia, claridad, coherencia y relevancia, para cada uno de los ítems. Para el análisis cuantitativo, se trabajó con el modelo de Hyrkäs et al. (2003), los estadísticos de Friedman, la W de Kendall y la V de Aiken. El análisis cualitativo se desarrolló siguiendo la metodología de la Teoría Fundamentada de Strauss y Corbin (2016). Los resultados evidencian óptimos niveles de calidad en el dominio específico de la variable medida y evaluada. En consecuencia, indican que el instrumento sometido a evaluación de jueces expertos presenta indicadores psicométricos coherentes y adecuados en relación con las variables que pretende indagar.

Palabras clave: validación de instrumento, juicio de expertos, bullying, cyberbullying, validez y confiabilidad, metodología de investigación, educación secundaria

Abstract

The objective of this article is to validate the items of a psychometric test on bullying and cyberbullying through the constitution of an expert panel. A mixed methodology with quantitative predominance was used based on the demonstration of four theoretical criteria (sufficiency, clarity, coherence and relevance) for each of the items. For the quantitative analysis, we worked with the model of Hyrkäs et al. (2003), Friedman's statistics, Kendall's W and Aiken's V. The qualitative analysis was developed following the methodology of Strauss and Corbin's Grounded Theory (2016). The results show optimal levels of quality in the specific domain of the variable measured and evaluated. Consequently, they indicate that the instrument submitted for evaluation by a panel of experts presents coherent and adequate psychometric indicators in relation to the variables it intends to investigate.

Keywords: instrument validation, expert panel, bullying, cyberbullying, validity and reliability, research methodology, middle education

*Correspondencia a: Cristián David Expósito. Primitivo de la Reta 522, Dpto K, Mendoza, Argentina. Tel: 54 (0261) 429-2563. E-mail: cdexposito@uncu.edu.ar

Nota de autores: El presente artículo se enmarca en el Proyecto de Investigación Plurianual N° 2419 – CONICET.

Cómo citar este artículo: Expósito, C. D., Marsollier, R. G., Difabio de Anglat, H., & Castro-Santander, A. (2023). Construcción y validación del Cuestionario de Acoso y Ciberacoso Escolar (CACE) mediante juicio de expertos. *Revista Evaluar*, 23(1), 61-79. Recuperado de <https://revistas.unc.edu.ar/index.php/revaluar>

Participaron en la edición de este artículo: Abigail Pérez, Ma. Eugenia Maiorana, Eugenia Barrionuevo, Pablo Carpintero, Agustina Mangieri, Florencia Ruiz, Mónica Serppe, Jorge Bruera.

Desarrollo

Los reportes de la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO, 2019) dejan en evidencia que casi un tercio de los chicos del planeta en edad escolar sufren de algún tipo de abuso relacionado con el acoso dentro de las instituciones educativas. Esta situación tiene un impacto negativo en el rendimiento académico y en la salud mental de las víctimas; a su vez, aumenta la ideación suicida y el riesgo de cometer suicidio (Cunha et al., 2023). Es inminente poner fin a esta situación de violencia si se pretende alcanzar los Objetivos de Desarrollo Sostenible para así garantizar una educación inclusiva, equitativa y de calidad con el fin de promover sociedades pacíficas e integradoras (Comisión Económica para América Latina [CEPAL], 2019).

Desde una perspectiva académica, se considera necesario contar con instrumentos e insumos científicos adecuados que sean lo más precisos posibles para abordar esta problemática en nuestros entornos regionales (Fernández et al., 2010). Por otra parte, la construcción de escalas contextualizadas surge ante la necesidad de contar con instrumentos ajustados a la realidad cultural y social que se espera conocer, como por ejemplo, la convivencia escolar (Ascorra-Costa et al., 2020).

Sin lugar a dudas, con el crecimiento exponencial de las tecnologías acaecido a lo largo de este último medio siglo, se ha generado una revolución en el campo de las ciencias. Tal es así que la información es la materia prima para el desarrollo social en la actualidad (Belloch, 2012) y es también el principal insumo de las ciencias en la generación del conocimiento (Pérez & Dressler, 2007). Sin embargo, este importante desarrollo tecnológico, que genera un gran impacto en la distribución de la información en el mundo, no siempre está acompañado de buenas intenciones.

Cada vez se complejiza más la tarea de discernir la información verdadera de la falsa y el uso que se hace de la misma de manera mediática y política a través de *fake news* masivos (Parra-Valero & Oliveira, 2018). Esta vorágine de la posverdad crea una atmósfera social nociva debido a que solo promueve desinformación (Novalbos, 1998).

La ciencia se caracteriza por seguir un protocolo denominado «método científico» que garantiza evidenciar un conocimiento nuevo (Otzen et al., 2017) y, de esta manera, evitar el error dentro de sus propios parámetros (Artigas, 1992). Es por esta razón que cuando el investigador se enfrenta a una determinada situación empírica, la tarea de identificar y conceptualizar dicha problemática en términos científicos (es decir, variables) implica un gran esfuerzo y una importante responsabilidad (Pies & Koslowski, 2011). Una vez que estas variables se logran definir y se pueden identificar los indicadores más adecuados para estudiarlas, surge la duda sobre qué tan confiable sería esta intervención. Este temor se basa en la posible incongruencia que pueda darse entre el dato obtenido mediante la aplicación del instrumento y la realidad en su propia complejidad. El instrumento debe brindar información conceptual y empírica de calidad (Matheus-Marín et al., 2017).

Existe una preocupación en el mundo erudito de las ciencias sociales por evidenciar un respaldo en el impacto de la validez de los resultados de las investigaciones (Hanney et al., 2017; Poppy, 2015; Solans-Domènech et al., 2019). Una de las alternativas más efectivas para demostrar la validación de una propuesta de investigación es el trabajo con jueces expertos (Cabero-Almenara et al., 2020). Desde el punto de vista científico, esta metodología se considera como una herramienta que brinda pertinencia y viabilidad a la propuesta de investigación (Pérez-Iribar et al., 2017). En este sentido, la consulta a un equipo de jueces expertos es un método heurístico con un importante

rigor científico, ya que su objetivo es alcanzar cierto nivel de consenso en cuestiones propias a su experticia, partiendo de aproximaciones cualitativas (González-Viera et al., 2018).

Por ello, el objetivo principal propuesto en este escrito es validar los ítems que conformarán una prueba psicométrica sobre *bullying* y *ciberbullying* a través de la constitución de un tribunal de expertos (Escobar-Pérez & Cuervo-Martínez, 2008).

Metodología

Diseño

La metodología utilizada es de carácter mixta con predominancia cuantitativa, con el objetivo de validar el contenido de un instrumento psicométrico mediante el análisis de un juicio de expertos. La validación de contenido implica identificar la coherencia que existe entre el ítem y la variable que pretende medir (Cohen & Swerdlik, 2001); es decir, el grado en que los reactivos que conforman el instrumento son “una muestra representativa del dominio del contenido que se mide” (Escrura-Mayaute, 1988, p. 105). El proceso de validación que se emplea es a través del juicio de expertos, que Escobar-Pérez y Cuervo-Martínez (2008) definen como “una opinión informada de personas con trayectoria en el tema, que son reconocidas por otros como expertos cualificados en este y que pueden dar información, evidencia, juicios y valoraciones” (p. 29).

Instrumento

Se diseñó una herramienta para medir los niveles de *bullying* y *ciberbullying* en escuelas de nivel medio denominado Cuestionario de Acoso y Ciberacoso Escolar (CACE). Este instrumento

estaba compuesto, en un primer momento, de 55 ítems y tenía por objeto medir ocho variables contenidas en tres niveles o dimensiones.

Dimensión 1: Tipos de violencia. Remite a la identificación de las diferentes formas que pueden adoptar las expresiones de acoso escolar (Ayala-Carrillo, 2015; Bacchini et al., 2014; Benish-Weisman et al., 2022).

a. *Violencia física:* es el acto de provocar un perjuicio intencionado a un compañero haciendo uso de la fuerza física de forma directa (golpeándolo, pateándolo, etc.) e indirecta (dañando o escondiendo pertenencias ajenas). Esta acción se vuelve indicador de *bullying* si se ocasiona de manera reiterada y repetitiva contra una víctima (Álvarez et al., 2006; Arellano, 2007; Olivera-Carhuaz & Yupanqui-Lorenzo, 2020).

b. *Violencia verbal:* es un acto agresivo mediante el uso de palabras o gestos con el fin de burlarse de un compañero u ofenderlo, siempre de manera intencional y reiterada en el tiempo. Este tipo de violencia puede caracterizarse por ser presencial, es decir, cara a cara, o de modo virtual, a través de las redes (Arias-Sandoval, 2009; Nieto et al., 2018; Oliva-Zárate, 2013).

c. *Violencia relacional:* este tipo de ofensa se caracteriza por producirse a nivel social entre estudiantes. Generalmente, el o los agresores pretenden extorsionar o excluir a un compañero de un equipo de trabajo o grupo social. Por otro lado, también se asocia a perjuicios en las relaciones con pares de modo permanente (Romero-Abrio et al., 2018; Ortega-Barón et al., 2017).

Dimensión 2: Modalidades de acoso. Esta dimensión se focaliza en el propio contexto donde se produce el acto violento y permite discriminar si esta provocación es presencial o virtual.

a. *Bullying* o acoso presencial: es el típico caso de violencia escolar, sea esta física, verbal o relacional de tipo directo, que tiene lugar en el predio de la institución educativa y alrededores

(Castro-Santander, 2017; García-Montañez & Amaury-Ascensio-Martínez, 2015).

b. *Cyberbullying* o acoso virtual: este tipo de ataque es el que se lleva a cabo mediante el uso de redes digitales e internet, y tiene por objeto provocar a la víctima de manera verbal (escrita, gestual o extorsiva) y/o relacional (directa o indirecta), de manera deliberada y casi permanente (Castro-Santander, 2013; Gasso et al., 2018).

Dimensión 3: Configuración de roles. Pretende identificar y caracterizar los diferentes papeles que asumen los estudiantes en determinadas situaciones de acoso escolar entre pares (Castro-Santander, 2007). En esta dimensión se pueden evidenciar tres roles bien definidos:

a. Víctima: es el estudiante que sufre el acto violento, ya sea este físico, verbal o relacional y que se produzca de manera presencial o virtual (Carrillo-Izquierdo, 2021; Martínez-Torres et al., 2014).

b. Acosador: es el agente activo que permanentemente está perturbando a su víctima de manera física, verbal o social (Puértolas-Jiménez & Montiel-Juan, 2017; Tejada et al., 2021).

c. Testigo: es el sujeto que reconoce u observa la escena de *bullying* entre compañeros como espectador pasivo o activo (Carola-Pérez et al., 2013; Ubieto-Pardo, 2021).

Participantes

En el juicio de expertos participaron diez profesionales eméritos en sus áreas de investigación específicas con formación académica de posgrado y que pertenecían a cinco países de Latinoamérica (cinco de Argentina, uno de Chile, uno de Perú, uno de México y uno de Uruguay). Se tomó en consideración para su selección los principios propuestos por Skjong y Wentworth (2000) de contar con experiencia en esta tarea,

buena predisposición y tiempo, experticia reconocida públicamente y objetividad profesional. Además, se tuvo en cuenta la cantidad de publicaciones en las áreas que se les solicitaba evaluar (McGartland et al., 2003).

Para la selección del número se tomó como antecedente el trabajo de Hyrkäs et al. (2003), quienes manifiestan que diez jueces pueden ofrecer una correcta y confiable evaluación de un instrumento psicométrico.

Guía de evaluación

Para que el instrumento fuera sometido a una evaluación rigurosa y congruente, se confeccionó una guía de evaluación compuesta de cinco pasos:

1. Presentación: en este apartado se le brinda a cada juez una breve introducción del instrumento y se le solicitan sus datos personales e institución de filiación. Además, se les informa sobre los objetivos de la evaluación, las peculiaridades más relevantes del instrumento y las características de la población a la que está dirigido.

2. Matriz de variables: considerando que el instrumento tiene una perspectiva multidimensional y, además, propone variables complejas, se elaboró un cuadro de doble entrada donde se discriminan los tres niveles o dimensiones del estudio y las variables y los indicadores identificados para la medición. Cada uno de estos elementos estuvo acompañado por una breve descripción y conceptualización para ayudar en el posterior análisis de los ítems del test (Escobar-Pérez & Cuervo-Martínez, 2008).

3. Instrumento: en esta etapa se presenta el instrumento completo, con su consigna de trabajo, indicadores o ítems y la referencia de la escala Likert de frecuencia con cinco opciones: (0 = *Nunca*, 1 = *Pocas veces al año*, 2 = *Algunas*

veces al mes, 3 = *Algunas veces a la semana*, 4 = *Todos los días*) (Sullivan & Artino, 2013).

4. Criterios de evaluación: a partir de este punto, se sigue el modelo propuesto por Escobar-Pérez y Cuervo-Martínez (2008), el cual analiza cuatro categorías intrínsecas del test (suficiencia, claridad, coherencia y relevancia) mediante una escala Likert de cuatro niveles (1 = *No cumple con el criterio*, 2 = *Bajo Nivel*, 3 = *Moderado nivel*, y 4 = *Alto nivel*). A su vez, como un elemento relevante para la unificación de criterios entre los jueces, se suma una breve explicación de cada una de las categorías a considerar y una definición de lo que significa cada uno de los niveles propuestos en la escala Likert.

5. Planilla de evaluación: este es el recurso más relevante de la evaluación, ya que está diseñado como herramienta de recolección de datos mixtos provistos por cada juez. En esta plantilla, los expertos vuelcan sus apreciaciones (siguiendo la escala Likert propuesta) sobre cada uno de los ítems del test y, además, se les solicita que proporcionen una observación o valoración cualitativa de los mismos. Una vez completada esta planilla, se le propone a cada evaluador que brinde una reflexión personal sobre todo el instrumento evaluado.

Procesamiento

Los datos obtenidos a partir de la evaluación por juicio de experto son de carácter cuantitativo y cualitativo, razón por la cual se prevé viabilizar una triangulación. Se comienza con el procesamiento descriptivo que proponen Hyrkäs et al. (2003), quienes sostienen que si el 80% de los expertos está de acuerdo en que un ítem es válido, puede incorporarse al instrumento. Posteriormente, se calcula el coeficiente de Friedman y Kendall para cada una de las cate-

gorías propuestas por Escobar-Pérez y Cuervo-Martínez (2008) en relación con las variables medidas por el CACE. Para finalizar el análisis cuantitativo, se aplica el estadístico V de Aiken con el objetivo de corroborar el coeficiente de validez del contenido (Penfield & Giacobbi, 2004).

A su vez, se efectúa un análisis cualitativo de las observaciones volcadas por los jueces en la planilla suministrada. Dicha información se analiza conjuntamente con los datos cuantitativos en cada ítem (Ibarra-Piza et al., 2018). Los criterios para la valoración cualitativa también están enfocados en la claridad del ítem, su redacción, redundancia y coherencia con la variable que pretende medir. La metodología que se emplea para este análisis está basada en la *Grounded Theory* (Teoría Fundamentada) de Strauss y Corbin (2016).

Resultados

Análisis descriptivo

Tras los resultados obtenidos del panel de expertos, en primer lugar, se llevó a cabo un análisis descriptivo. En este caso, se promediaron los puntajes otorgados a cada ítem en particular en una escala de 1 a 4 puntos que evalúa suficiencia, claridad, coherencia y relevancia (Escobar-Pérez & Cuervo-Martínez, 2008). En términos generales, los puntajes otorgados por los jueces son aceptables para la mayoría de los ítems. La valoración más baja corresponde a los ítems 45 (2.70), 28 (2.75), 47 (2.87), 10 (2.90), 12 (2.90), 24 (2.95) e ítem 46 (2.98); mientras que los ítems 25, 32, 36, 40 y 42 obtuvieron un puntaje perfecto (4.00).

Siguiendo el modelo de Hyrkäs et al. (2003), se analiza la puntuación que cada juez otorga a cada ítem en suficiencia, claridad, coherencia y relevancia con el objetivo de identificar cuáles ítems pueden ser descartados. El planteo requiere que el

Tabla 1
Puntajes promedio por ítem.

Ítem	Juez 1	Juez 2	Juez 3	Juez 4	Juez 5	Juez 6	Juez 7	Juez 8	Juez 9	Juez 10	%
Ítem 08	3	3.25	3.50	3.50	3.25	3.25	2.50	3.25	3.50	3.25	90 %
Ítem 10	3	2.75	3.50	3	3	3.75	3	2.25	2.25	2.50	60 %
Ítem 12	2.25	3.50	3	3	3.25	4	2	2.75	2.50	2.75	50 %
Ítem 13	2.50	3.25	3.25	3.25	3.25	3.50	3	3	3.25	3.25	90 %
Ítem 23	3.50	3.75	3.25	3.50	3.75	3.50	2.50	2.75	3.25	3	80 %
Ítem 24	3	3.25	3.50	3.50	3	3.25	2	1.75	3.25	3.25	80 %
Ítem 28	2.75	2.50	3.25	3.25	3	3.25	2	3	2.25	2.25	50 %
Ítem 41	4	4	4	4	4	4	4	4	2.50	2.25	80 %
Ítem 44	3	3	3	3	3	2.75	2.75	2.50	3.25	3	70 %
Ítem 45	2.50	2.75	2.50	3	2.50	2.75	2.50	2.50	3	3	30 %
Ítem 46	3.25	3.25	3	3	3	3.25	2	3	3.25	3.50	90 %
Ítem 47	2.75	3	3	3	3	3	2	2.75	3	3.25	70 %
Ítem 50	3.50	3	3.25	3.25	2.75	3.50	1.75	3.25	3	3.50	80 %
Ítem 51	2	2.75	3.50	3	3.50	3.50	1.25	3.25	3	3.75	70 %
Ítem 52	3.50	3	3.50	3.50	3.50	3.50	2.75	2.75	3.50	3.50	80 %
Ítem 53	3	3.25	3	3	2.75	3.50	2.50	3.50	3	3.25	80 %
Ítem 55	2.50	3	3.25	3.50	3.25	3.50	3	3	3.25	3.25	90 %

80% de los evaluadores considere válido el ítem mediante la asignación de un puntaje promedio superior o igual a 3 puntos (Válido ≥ 3).

En orden a la brevedad, solo se consignan en la Tabla 1 los 17 ítems que recibieron al menos una puntuación inferior a 3 de parte de alguno de los jueces.

De estos 17 ítems consignados en la Tabla 1, podemos observar que siete de ellos (ítems 10, 12, 28, 44, 45, 47 y 51) no cumplen con los criterios propuestos para su validación, ya que el 20% o más de los jueces asignaron puntuaciones por debajo de 3. Se considera que esta valoración es insuficiente debido a que muchos ítems con puntajes promedio por encima de 3 presentan dificultades serias en algunos de los criterios evaluados. Es por esta razón que se efectúa una lectura en mayor profundidad mediante el análisis cualitativo de estos 17 ítems. Cabe destacar que son los ítems que obtuvieron menor puntuación promedio en general.

Análisis cualitativo

Con el objetivo de comprender los motivos por los cuales los jueces puntuaron estos 17 ítems con valores inferiores, se realizó una lectura cualitativa de las observaciones para etiquetar, codificar (de manera abierta, axial y selectiva) y categorizar las opiniones (Soneira, 2006; Vasilachis de Gialdino, 2006), siguiendo la metodología de la Teoría Fundamentada (Strauss & Corbin, 2016).

Mediante este análisis se pudo identificar un total de 47 falencias viables de ser categorizadas y señaladas por los diez jueces. Como consecuencia, surgieron tres posibles categorías explicativas, a saber, 1) Inaceptable, 2) Redundante y 3) Modificable. Las categorías 1 y 2 implican la eliminación del ítem, mientras que la categoría 3 demanda reformulación en la redacción. A continuación, se presentan algunos ejemplos sobre estas opiniones:

Tabla 2

Dictamen en relación con la categorización de cada ítem en conflicto, el puntaje promedio y el porcentaje de aceptación.

Ítem	Categorías	Puntaje Promedio	% de aceptación	Dictamen
Ítem 08	Modificable	3.23	90 %	Ajustar sintaxis
Ítem 10	Inaceptable	2.90	60 %	Eliminar
Ítem 12	Inaceptable	2.90	50 %	Eliminar
Ítem 13	Modificable	3.08	90 %	Ajustar sintaxis
Ítem 23	Modificable	3.28	80 %	Ajustar sintaxis
Ítem 24	Redundante con ítem 23	2.95	80 %	Eliminar
Ítem 28	Inaceptable	2.75	50 %	Eliminar
Ítem 41	Redundante con ítem 42	3.68	80 %	Eliminar
Ítem 44	Redundante con ítem 34	3.15	70 %	Eliminar
Ítem 45	Inaceptable	2.70	30 %	Eliminar
Ítem 46	Redundante con ítem 43	2.98	90 %	Eliminar
Ítem 47	Redundante con ítem 42	2.88	70 %	Eliminar
Ítem 50	Modificable	3.05	80 %	Ajustar sintaxis
Ítem 51	Inaceptable	2.93	70 %	Eliminar
Ítem 52	Modificable	3.30	80 %	Ajustar sintaxis
Ítem 53	Redundante con ítem 46	3.08	80 %	Eliminar
Ítem 55	Modificable	3.15	90 %	Ajustar sintaxis

- “Podría ser claro en que se trata de golpear/trompadas como un acto de *bullying*. Es muy relevante el ítem, pero con esa reconfiguración. Otra opción es eliminarlo, ya que el ítem 11 incluye «(...) pegándole a otros chicos»”. (Juez 1 en referencia al ítem 10).

- “Muchas veces no es con palabras groseras. Puede darse con adjetivos que molestan, y son aplicados por la molestia que muestra el acosado y no por preferencia del acosador, ni por el significado del término usado. Otras veces el insulto se aplica a través de tonos de voz, sonidos, gestos, etc.”. (Juez 2 en referencia al ítem 13).

- “Es reiterativo con el ítem 34; podrían integrarse”. (Juez 8 en referencia al ítem 44).

- “Creo que sería mejor utilizar la conjugación «ridiculizan» en vez de «ridiculicen»”. (Juez 6 en referencia al ítem 55).

En la Tabla 2 se presentan las opiniones de los jueces categorizadas para los 17 ítems en conflicto junto con el puntaje promedio obtenido en el

análisis descriptivo y el porcentaje de jueces que aceptaron el ítem.

Según la lectura de las opiniones de los jueces con relación a estos ítems analizados y el puntaje promedio asignado, se puede considerar la eliminación de 11 ítems (10, 12, 24, 28, 41, 44, 45, 46, 48, 51 y 53) y la reformulación de los seis ítems restantes (08, 13, 23, 50, 52, y 55).

En consecuencia, la distribución de ítems quedaría conformada de la siguiente manera según la puntuación promedio otorgada por los jueces:

En la Figura 1 se puede observar en color rosado los 11 ítems sometidos a eliminación; en verde, aquellos ítems que se recomendó modificar; y en azul, los que se pueden mantener tal cual están presentados.

Con base en esta información, se analizan todas las puntuaciones de cada ítem mediante los estadísticos de Friedman y Kendall (Escobar-Pérez & Cuervo-Martínez, 2008).

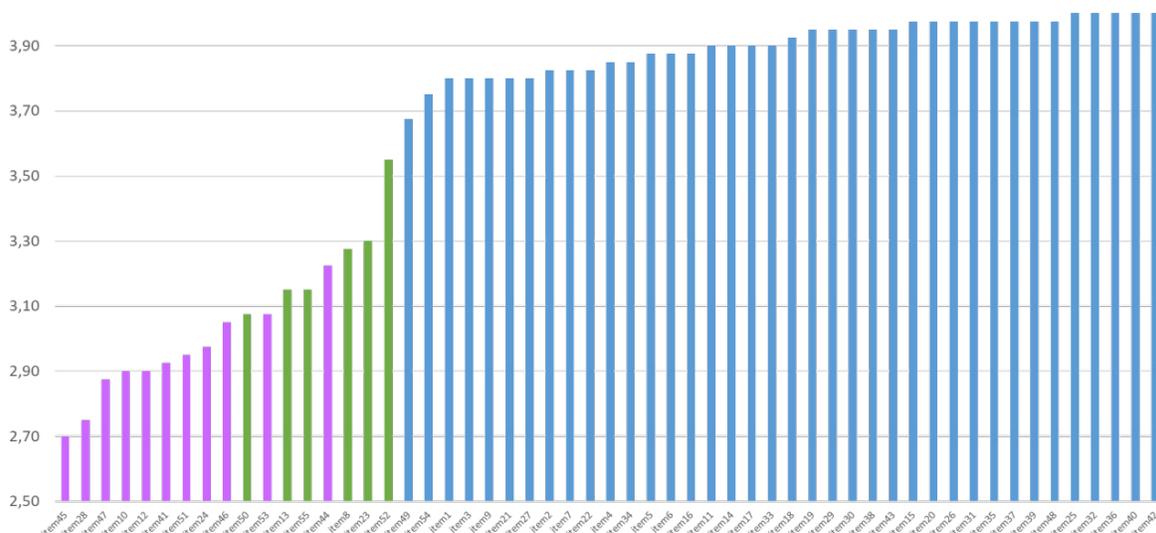


Figura 1

Puntajes promedio otorgados por los jueces a cada ítem.

Coefficiente de Friedman y Kendall

El análisis de validez de contenido es un proceso multimetódico que demanda un aporte cuantitativo y cualitativo de forma simultánea (Ibarra-Piza et al., 2018). Cabe destacar que este tipo de estadístico (Friedman y Kendall) es pertinente, ya que la prueba de Kolmogorov-Smirnov establece que no existe una distribución normal en la muestra ($p < .05$) (Berger & Zhou, 2014).

Las pruebas de Friedman y Kendall se utilizan para determinar la concordancia entre las respuestas de los jueces (Granato et al., 2014; Lugo & Ferrer, 2019). El estadístico de Friedman permite rechazar la hipótesis que plantea la inexistencia de diferencias significativas entre medianas de las distribuciones de probabilidad de las variables asociadas a cada ítem si el p valor es menor a .05. De esta manera se pone en evidencia que cada ítem del test aporta información relevante para la variable que pretende medir (Baena-González et al., 2021). La W de Kendall suma la fuerza de concordancia entre los jueces, la cual aumenta cuando se acerca a 1 (Escobar-Pérez & Cuervo-Martínez, 2008).

Con base en el modelo de Escobar-Pérez y Cuervo-Martínez (2008), se analizan los valores aportados por los jueces (entre 1 y 4) en relación con los criterios de claridad, suficiencia, coherencia y relevancia para los 17 ítems en conflicto.

El resultado obtenido mediante el estadístico de Friedman para los ítems en conflicto indica que existen niveles significativos de acuerdo entre los jueces, excepto para los reactivos 24 y 41, cuyos p valores están por encima de .05.

Con respecto al nivel de concordancia entre los jueces obtenido mediante el estadístico de Kendall (Dorantes-Nova et al., 2016), el nivel más elevado se presenta para el ítem 53 ($W = .903$) y los más bajos corresponden a los ítems 24 ($W = .058$) y 41 ($W = .171$). Sin embargo, es necesario tener en cuenta que estos ítems (24 y 41) fueron cuestionados por redundancia con otros reactivos; por esta razón se mantiene la sentencia de eliminación (Ver Tabla 3).

A partir de los tres análisis presentados, se toma la decisión de eliminar los 11 ítems cuestionados. Para asegurar la claridad, suficiencia, coherencia y relevancia se someten los ítems restantes a la prueba V de Aiken.

Tabla 3

Dictamen en relación con la concordancia de cada ítem en conflicto (Friedman y Kendall).

Ítem	χ^2	<i>p</i> valor (Friedman)	W de Kendall	Categoría	Dictamen
Ítem 08	12.706	< .05	.424	Modificable	Ajustar sintaxis
Ítem 10	14.301	< .05	.477	Inaceptable	Eliminar
Ítem 12	9.828	< .05	.328	Inaceptable	Eliminar
Ítem 13	12.791	< .05	.426	Modificable	Ajustar sintaxis
Ítem 23	19.227	< .05	.641	Modificable	Ajustar sintaxis
Ítem 24	1.7462	> .05	.058	Redundante con ítem 23	Eliminar
Ítem 28	16.291	< .05	.543	Inaceptable	Eliminar
Ítem 41	5.117	> .05	.171	Redundante con ítem 42	Eliminar
Ítem 44	17.625	< .05	.587	Redundante con ítem 34	Eliminar
Ítem 45	23.561	< .05	.785	Inaceptable	Eliminar
Ítem 46	19.107	< .05	.637	Redundante con ítem 43	Eliminar
Ítem 47	28.814	< .05	.960	Redundante con ítem 42	Eliminar
Ítem 50	22.360	< .05	.745	Modificable	Ajustar sintaxis
Ítem 51	16.625	< .05	.554	Inaceptable	Eliminar
Ítem 52	27.000	< .05	.900	Modificable	Ajustar sintaxis
Ítem 53	27.075	< .05	.903	Redundante con ítem 46	Eliminar
Ítem 55	21.835	< .05	.728	Modificable	Ajustar sintaxis

Estadístico V de Aiken

La validez de un instrumento está íntimamente relacionada con el grado de medición del test. Para ello, se “calcula la razón de un dato obtenido sobre la suma máxima de la diferencia de los valores posibles” (Eскурra-Mayaute, 1988, p. 107).

$$V = \frac{S}{(n(c - 1))}$$

Referencia:

- S: Sumatoria de puntaje
- n: Número de jueces
- c: Valor máximo de la escala (4 en este caso)

Este coeficiente de validez de contenido de Aiken permite obtener un valor que oscila entre 0 y 1, donde 0 indica que no existe validez y 1 implica máxima validez del contenido. Además, po-

see una significancia estadística y un intervalo de confianza del 95% (Penfield & Giacobbi, 2004).

Los ítems retenidos alcanzan valores superiores a .90 en todas las categorías; sin embargo, los ítems que demandan reformulación presentaron valores por debajo de .80 en algunos casos. En orden a la brevedad, solo se presentan los 6 ítems conflictivos recodificados.

En la Tabla 4 se puede observar que los puntos críticos de estos ítems son la claridad (08, 11, 21 y 41) y la relevancia (42 y 44). Respecto de la suficiencia y la coherencia, presentan valores aceptables por encima de .80 (Ibarra-Piza et al., 2018).

Los resultados del análisis *V* de Aiken presentaron valores superiores a .90 para todas las categorías y el valor de intervalo de confianza presentó valores por encima de .80. Esto evidencia la escasa dispersión de los valores del coeficiente (*V* de Aiken) y denota la existencia de diferencias significativas ($p < .05$) en todas las categorías (Eскурra-Mayaute, 1988). (Ver Tabla 5).

Tabla 4Validez de contenido mediante *V* de Aiken por ítem conflictivo.

Ítem recodificado	Suficiencia	Claridad	Coherencia	Relevancia
Ítem 08 (08*)	.88	.58	.85	.93
Ítem 11 (13*)	.95	.60	.80	.80
Ítem 21 (23*)	.90	.50	.95	.93
Ítem 41 (50*)	.85	.40	.95	.88
Ítem 42 (52*)	.90	.98	1.0	.43
Ítem 44 (55*)	.85	.98	.93	.40

Nota. *Codificación anterior.**Tabla 5**Validez de contenido mediante *V* de Aiken y los intervalos de confianza al 95%.

Categorías	V de Aiken	Sig. (bilateral)	95 % de intervalo de confianza de la diferencia	
			Inferior	Superior
Suficiencia	.972	.000 (< .05)	.959	.985
Claridad	.911	.000 (< .05)	.870	.951
Coherencia	.972	.000 (< .05)	.959	.985
Relevancia	.948	.000 (< .05)	.910	.986

Nota. Valor de prueba = 0

Discusión

Todo instrumento de recolección de datos que se precie de efectivo debe ser evaluado por expertos en la cuestión que aprueben o refuten su diseño (Carvajal et al., 2011). Esto es lo que permite brindarle solidez en el diseño y en la confiabilidad para luego llevar a cabo su aplicación (Moreira-Mora, 2021). Además, se debe comprobar si cumple con los criterios métricos de validez, confiabilidad y calidad (Hernández-Sampieri & Mendoza-Torres, 2018), sin perder de vista la viabilidad y factibilidad de aplicación (Lugo & Ferrer, 2019).

El CACE es un instrumento que demanda una serie de consideraciones específicas al momento de su evaluación debido a la complejidad del tema que aborda. Es por esta razón que se concuerda con la postura de aquellos autores que sostienen la necesidad de realizar un análisis multimetódico o mixto al momento de evaluar

este instrumento como constructo psicométrico (Bulger & Housner, 2007; Haynes et al., 1995; Ibarra-Piza et al., 2018; Penfield & Giacobbi, 2004; Sireci, 1998).

El proceso de evaluación está basado en el modelo estructural de Escobar-Pérez y Cuervo-Martínez (2008), ya que se considera adecuado para brindar a los jueces un método claro y preciso de calificación mixta de la información suministrada (Galicia-Alarcón et al., 2017; Carbajal & Tamari, 2017; Juárez-Hernández & Tobón, 2018).

Si bien el proceso de validación mediante juicio de expertos es un sistema recomendado y muy efectivo, una de sus limitaciones más relevantes refiere a que dicha validación se sustenta en las opiniones de los expertos y no en datos extraídos del campo (Bruna-Jofré et al., 2020; Kane, 2013; Pedrosa et al., 2014). En consecuencia, el instrumento debe ser sometido en un futuro próximo a una serie de pruebas denominadas “piloto” para poder demostrar la validez, objeti-

vidad y confiabilidad de los indicadores mediante estadísticos más imparciales.

Conclusión

Los datos analizados permitieron identificar falencias significativas en 17 ítems según el criterio de los 10 expertos. Once de estas falencias eran irreversibles, mientras que seis reactivos fueron reescritos y ajustados a los criterios de sintaxis y dimensiones especificadas.

Gracias a la colaboración de los jueces, se visibilizaron seis ítems que generaban redundancia en la información; además, surgieron a la luz conceptos ambiguos y poco claros en la redacción de los reactivos.

Sin embargo, los resultados de las diferentes pruebas estadísticas de Kendall y Friedman y de Aiken a las que se sometieron las respuestas de los expertos pusieron en relieve la validez del contenido del instrumento y la calidad en el dominio específico de la variable que evalúa. En consecuencia, estos indican que el instrumento sometido a evaluación de jueces expertos presenta cualidades psicométricas coherentes y adecuadas en relación con las variables que pretenden indagar. Por consiguiente, en vías a su validación empírica, la versión que será aplicada en una prueba piloto se constituye de 44 reactivos.

Referencias

Álvarez, L., Álvarez, D., González-Castro, P., Núñez, J., & González-Pianda, J. (2006). Evaluación de los comportamientos violentos en los centros educativos. *Psicothema*, 18(4), 686-695. <https://www.psicothema.com>

Arellano, N. (2007). La violencia escolar y la provención del conflicto. *Orbis: Revista de Ciencias Humanas*,

3(7), 23-45. <http://www.revistaorbis.org>

- Arias-Sandoval, L. (2009). Una mirada al fenómeno de la violencia escolar en Costa Rica. *Revista Electrónica Educare*, 13(1), 41-51. <https://doi.org/10.15359/ree.13-1.4>
- Artigas, M. (1992). Conocimiento humano, fiabilidad y fiabilidad. *Anuario Filosófico*, 25(2), 277-294. <https://doi.org/10.15581/009.25.29954>
- Ascorra-Costa, P., Cárdenas-Mancilla, K., & Álvarez-Figueroa, F. (2020). Gestión de la convivencia escolar a nivel intermedio: Diseño y validación de una escala. *Revista Evaluar*, 20(3), 1-19. <https://doi.org/10.35670/1667-4545.V20.N3.31700>
- Ayala-Carrillo, M. del R. (2015). Violencia escolar: Un problema complejo. *Ra Ximhai*, 11(4). <http://www.raximhai.com.mx/Portal>
- Bacchini, D., Affuso, G., & Aquilar, S. (2014). Multiple forms and settings of exposure to violence and values. *Journal of Interpersonal Violence*, 30(17), 3065-3088. <https://doi.org/10.1177/0886260514554421>
- Baena-González, R., García-Tascón, M., Chavarría-Ortiz, C., Maciá, M. J., & Gallardo, A. M. (2021). Estudio preliminar para el diseño y validación de un cuestionario sobre la percepción de la seguridad según el sistema antivuelco de las porterías en eventos deportivos. *Cultura, Ciencia y Deporte*, 16(50), 583-592. <https://doi.org/10.12800/ccd.v16i50.1613>
- Belloch, C. (2012). Las tecnologías de la información y comunicación en el aprendizaje. En Recursos tecnológicos en educación y logopedia. <https://www.uv.es/belloch>
- Benish-Weisman, M., Oreg, S., & Berson, Y. (2022). The contribution of peer values to children's values and behavior. *Personality and Social Psychology Bulletin*, 48(6), 844-864. <https://doi.org/10.1177/01461672211020193>
- Berger, V. W., & Zhou, Y. (2014). Kolmogorov-Smirnov Test: Overview. *Wiley StatsRef: Statistics Reference Online*. <https://doi.org/10.1002/9781118445112.STAT06558>
- Bruna-Jofré, C., Espinoza-Parcet, C., Fernández-Branada,

- C., Sánchez-Bizama, J., & Díaz-Bormann, E. (21, 22 y 23 de octubre de 2020). Validez de contenido de instrumento para evaluar percepción de estudiantes sobre intervenciones pedagógicas en aulas universitarias. En COINCOM, Libro de Congreso: *IX Congreso Internacional sobre Competencias y Educación COINCOM2020 "Modernización de la Educación en la Era Digital"*. Universidad de Concepción Chile. <http://memoriascimted.com>
- Bulger, S. M., & Housner, L. D. (2007). Modified Delphi investigation of exercise science in physical education teacher education. *Journal of Teaching in Physical Education*, 26(1), 57-80. <https://doi.org/10.1123/JTPE.26.1.57>
- Cabero-Almenara, J., Romero-Tena, R., & Palacios-Rodríguez, A. (2020). Evaluation of teacher digital competence frameworks through expert judgement: The use of the Expert Competence Coefficient. *Journal of New Approaches in Educational Research (NAER Journal)*, 9(2), 275-293. <https://doi.org/10.7821/naer.2020.7.578>
- Carbajal, A. B., & Tamari, K. (2017). Validez de contenido y confiabilidad inter-observadores de Escala Integral Calidad de Vida. *Revista de Psicología (PUCP)*, 35(2), 641-666. <https://doi.org/10.18800/PSICO.201702.009>
- Carola-Pérez, J., Astudillo, J., Jorge-Varela, T., & Felipe-Lecannelier, A. (2013). Evaluación de la efectividad del Programa Vínculos para la prevención e intervención del bullying en Santiago de Chile. *Psicología Escolar e Educativa*, 17(1), 163-172. <https://doi.org/10.1590/S1413-85572013000100017>
- Carrillo-Izquierdo, A. (2021). La víctima en el bullying. En M. E. Nieto-Cabrera (Ed.), *Víctimas sociales y víctimas de delitos: La promoción personal y social a través de la intervención* (pp. 211-226). <https://www.torrossa.com>
- Carvajal, A., Centeno, C., Watson, R., Martínez, M., & Sanz-Rubiales, Á. (2011). ¿Cómo validar un instrumento de medida de la salud? *Anales del Sistema Sanitario de Navarra*, 34(1), 63-72. <https://recyt.fecyt.es/index.php/ASSN/index>
- Castro-Santander, A. (2007). *Violencia silenciosa en la escuela. Dinámica del acoso escolar y laboral*. Bonum.
- Castro-Santander, A. (2013). Formar para la ciberconvivencia Internet y prevención del ciberbullying. *Revista Integra Educativa*, 6(2), 49-70. http://www.scielo.org.bo/scielo.php?script=sci_serial&pid=1997-4043&lng=es&nrm=iso
- Castro-Santander, A. (2017). *Desaprender la violencia: Un nuevo desafío educativo*. Bonum.
- Cohen, R. J., & Swerdlik, M. E. (2001). *Pruebas y evaluación psicológicas: Introducción a las pruebas y a la medición* (4ª ed.). McGraw-Hill.
- Comisión Económica para América Latina. (2019). *La agenda 2030 y los objetivos de desarrollo sostenible: Una oportunidad para América Latina y el Caribe. Objetivos, metas e indicadores mundiales*. CEPAL. <https://www.cepal.org/es>
- Cunha, F., Hu, Q., Xia, Y., & Zhao, N. (2023). *Reducing bullying: Evidence from a parental involvement program on empathy education*. NBER Working Paper No. 30827. <https://doi.org/10.3386/W30827>
- Dorantes-Nova, J. A., Hernández-Mosqueda, J. S., & Tobón-Tobón, S. (2016). Juicio de expertos para la validación de un instrumento de medición del Síndrome de Burnout en la docencia. *Ra Ximhai: Revista Científica de Sociedad, Cultura y Desarrollo Sostenible*, 12(6), 327-346. <http://www.raximhai.com.mx/Portal>
- Escobar-Pérez, J., & Cuervo-Martínez, Á. (2008). Validez de contenido y juicio de expertos: Una aproximación a su utilización. *Avances en Medición*, 6, 27-36. https://www.humanas.unal.edu.co/lab_psicometria/revista-avances-en-medicion
- Escurra-Mayaute, L. M. (1988). Cuantificación de la validez de contenido por criterio de jueces. *Revista de Psicología*, 6(1-2), 103-111. <https://doi.org/10.18800/PSICO.198801-02.008>
- Fernández, A., Pérez, E., Alderete, A. M., Richaud, C., & Fernández-Liporace, M. F. (2010). ¿Construir o adaptar tests psicológicos? Diferentes respuestas a

- una cuestión controvertida. *Revista Evaluar*, 10(1), 60-74. <https://doi.org/10.35670/1667-4545.V10.N1.459>
- Galicia-Alarcón, L. A., Balderrama-Trápaga, J. A., & Navarro, R. E. (2017). Validez de contenido por juicio de expertos: Propuesta de una herramienta virtual. *Apertura*, 9(2), 42-53. <https://doi.org/10.32870/Ap.v9n2.993>
- García-Montañez, M., & Amaury-Ascensio-Martínez, C. (2015). Bullying y violencia escolar: Diferencias, similitudes, actores, consecuencias y origen. *Revista Intercontinental de Psicología y Educación*, 17(2), 9-38. <https://dialnet.unirioja.es/servlet/revista?codigo=17329>
- Gasso, A. M., Fernández-Cruz, V., Montiel-Juan, I., & Agustina, J. R. (2018). Violencia escolar a través de medios digitales: Del bullying al cyberbullying. *Educación y Futuro: Revista de Investigación Aplicada y Experiencias Educativas*, 38, 57-82. <http://hdl.handle.net/20.500.12328/1526>
- González-Viera, D., Socorro-Quesada, M., Gómez-Ravelo, I., de las Cuevas-Milán, H. R., & Fernández de Castro-Fabré, A. (2018). Indicators for the evaluation of impact of the technological administration in rice tillage. *Revista Ciencias Técnicas Agropecuarias*, 39(1), 99-109. <http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=129551922&lang=es&site=ehost-live>
- Granato, D., de Araújo-Calado, V. M., & Jarvis, B. (2014). Observations on the use of statistical methods in food science and technology. *Food Research International*, 55, 137-149. <https://doi.org/10.1016/J.FOODRES.2013.10.024>
- Hanney, S., Greenhalgh, T., Blatch-Jones, A., Glover, M., & Raftery, J. (2017). The impact on healthcare, policy and practice from 36 multi-project research programmes: Findings from two reviews. *Health Research Policy and Systems*, 15, Article 26. <https://doi.org/10.1186/s12961-017-0191-y>
- Haynes, S. N., Richard, D. C. S., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment*, 7(3), 238-247. <https://doi.org/10.1037/1040-3590.7.3.238>
- Hernández-Sampieri, R., & Mendoza-Torres, C. (2018). *Metodología de la investigación: Las rutas cuantitativa, cualitativa y mixta*. Mc Graw Hill.
- Hyrkäs, K., Appelqvist-Schmidlechner, K., & Oksa, L. (2003). Validating an instrument for clinical supervision using an expert panel. *International Journal of Nursing Studies*, 40(6), 619-625. [https://doi.org/10.1016/S0020-7489\(03\)00036-1](https://doi.org/10.1016/S0020-7489(03)00036-1)
- Ibarra-Piza, S., Segredo-Santamaria, S., Juárez-Hernandez, L., & Tobón, S. (2018). Estudio de validez de contenido y confiabilidad de un instrumento para evaluar la metodología socioformativa en el diseño de cursos. *Revista Espacios*, 39(53). <https://www.revistaespacios.com>
- Juárez-Hernández, L. G., & Tobón, S. (2018). Análisis de los elementos implícitos en la validación de contenido de un instrumento de investigación. *Revista Espacios*, 39(53). <https://www.revistaespacios.com>
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1-73. <https://doi.org/10.1111/JEDM.12000>
- Lugo, N. T., & Ferrer, R. L. (2019). Validación de un cuestionario sobre factores de riesgo para defectos congénitos. *Revista Cubana de Investigaciones Biomédicas*, 38(4). <http://www.revibiomedica.sld.cu/index.php>
- Martínez-Torres, J., Ganem, A. E., Contreras, M. M., Leal, E. K., Soto, M., & Fernández, J. (2014). Prevalencia y factores de riesgo para ser víctima de bullying en escolares de 8 a 12 años de edad en una escuela pública. *Revista Chilena de Terapia Ocupacional*, 14(1), 81-88. <https://doi.org/10.5354/0719-5346.2014.32392>
- Matheus-Marín, A. C., Romero-López, R., & Parroquín-Amaya, P. C. (2017). Validación por expertos de un instrumento para la identificación de habilidades y competencias de un profesional en el área de logística. *CULCyT: Cultura Científica y Tecnológica*, 63(14-Especial), 227-238. <https://revistas.uacj.mx/>

ojs/index.php/culcyt/index

- McGartland, D., Berg-Weger, M., Tebb, S., Lee, E., & Rauch, S. (2003). Objectifying content validity: Conducting a content validity study in social work research. *Social Work Research, 27*(2), 94-104. <https://doi.org/10.1093/swr/27.2.94>
- Moreira-Mora, T. E. (2021). Propiedades psicométricas de una prueba de admisión universitaria. *Revista Evaluar, 21*(1), 73-93. <https://doi.org/10.35670/1667-4545.V21.N1.32833>
- Nieto, B., Portela, I., López, E., & Domínguez, V. (2018). Verbal violence in students of compulsory secondary education. *European Journal of Investigation in Health, Psychology and Education, 8*(1), 5-14. <https://www.mdpi.com/journal/ejihpe>
- Novalbos, L. (1998). Los efectos de “La guerra de los mundos” en la audiencia. Factores desencadenantes del “pánico radiofónico”. *Semiósfera: Humanidades-Tecnologías, 8*, 119-137. <https://e-archivo.uc3m.es/handle/10016/6600>
- Oliva-Zárate, L. (2013). La violencia escolar desde la infancia hasta la juventud. *Revista Costarricense de Psicología, 32*(2), 137-154. <http://rcps-cr.org>
- Olivera-Carhuaz, E., & Yupanqui-Lorenzo, D. (2020). Violencia escolar y funcionalidad familiar en adolescentes con riesgo de deserción escolar. *Revista Científica de la UCSA, 7*(3), 3-13. <http://revista-ucsa-ct.edu.py/ojs/index.php/ucsa/index>
- Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. (2019). *Behind the numbers: Ending school violence and bullying*. UNESCO. <https://unesdoc.unesco.org>
- Ortega-Barón, J., Buelga, S., Cava, M. J., & Torralba, E. (2017). Violencia escolar y actitud hacia la autoridad de estudiantes agresores de cyberbullying. *Revista de Psicodidáctica, 22*(1), 23-28. [https://doi.org/10.1016/S1136-1034\(17\)30040-0](https://doi.org/10.1016/S1136-1034(17)30040-0)
- Otzen, T., Manterola, C., Rodríguez-Núñez, I., & García-Domínguez, M. (2017). La necesidad de aplicar el método científico en investigación clínica: Problemas, beneficios y factibilidad del desarrollo de protocolos de investigación. *International Journal of Morphology, 35*(3), 1031-1036. <https://doi.org/10.4067/S0717-95022017000300035>
- Parra-Valero, P., & Oliveira, L. (2018). Fake news: Una revisión sistemática de la literatura. *Observatorio (OBS*)*, *Special Issue*, 54-78. <https://doi.org/10.15847/obsOBS12520181374>
- Pedrosa, I., Suárez-Álvarez, J., & García-Cueto, E. (2014). Evidencias sobre la validez de contenido: Avances teóricos y métodos para su estimación. *Acción Psicológica, 10*(2), 3-18. <https://doi.org/10.5944/ap.10.2.11820>
- Penfield, R. D., & 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. https://doi.org/10.1207/S15327841MPEE0804_3
- Pérez, D., & Dressler, M. (2007). Tecnologías de la información para la gestión del conocimiento. *Intangible Capital, 3*(1), 31-59. <https://www.intangiblecapital.org/index.php/ic>
- Pérez-Iribar, G., Beleño-Fuentes, M., Nuñez-Peña, C. R., & Orquera-Cadena, M. (2017). Valoración del resultado científico de la investigación. Una experiencia desde la aplicación del criterio de experto. *Olimpia: Publicación Científica de la Facultad de Cultura Física de la Universidad de Granma, 14*(46), 154-168. <https://revistas.udg.co.cu/index.php/olimpia/index>
- Pies, I., & Koslowski, P. (Eds.). (2011). *Corporate citizenship and new governance*. Springer Netherlands. <https://doi.org/10.1007/978-94-007-1661-2>
- Poppy, G. (2015). Science must prepare for impact. *Nature, 526*(7571), 7. <https://doi.org/10.1038/526007a>
- Puértolas-Jiménez, A., & Montiel-Juan, I. (2017). Bullying en la educación secundaria: Una revisión sobre las características de las víctimas y las víctimas-acosadores y las consecuencias de su victimización. *Revista de Victimología, 5*, 85-128. <http://www.huygens.es/journals/index.php/revista-de-victimologia/index>

- Romero-Abrio, A., Musitu, G., Callejas-Jerónimo, J. E., Sánchez-Sosa, J. C., & Villarreal-González, M. E. (2018). Factores predictores de la violencia relacional en la adolescencia. *Liberabit*, 24(1), 29-43. <https://doi.org/10.24265/LIBERABIT.2018.V24N1.03>
- Sireci, S. G. (1998). The construct of content validity. *Social Indicators Research*, 45(1), 83-117. <https://doi.org/10.1023/A:1006985528729>
- Skjong, R., & Wentworth, B. (2000). Expert judgement and risk perception. *International Society of Offshore and Polar Engineers*, 6, 537-544. <https://onepetro.org/ISOPEIOPEC/proceedings-abstract/ISOPE01/All-ISOPE01/8186>
- Solans-Domènech, M., Pons, J. M., Adam, P., Grau, J., & Aymerich, M. (2019). Development and validation of a questionnaire to measure research impact. *Research Evaluation*, 28(3), 253-262. <https://doi.org/10.1093/RESEVAL/RVZ007>
- Soneira, A. (2006). La teoría fundamentada en los datos (Grounded Theory) de Glaser y Strauss. En I. Vasilachis de Gialdino (Ed.), *Estrategias de Investigación Cualitativa* (pp. 153-173). Gedisa.
- Strauss, A., & Corbin, J. (2016). *Bases de la investigación cualitativa: Técnicas y procedimientos para desarrollar la teoría fundamentada* (2^{da} ed.). Editorial Universidad de Antioquia.
- Sullivan, G. M., & Artino, A. R. (2013). Analyzing and interpreting data from likert-type scales. *Journal of Graduate Medical Education*, 5(4), 541-542. <https://doi.org/10.4300/JGME-5-4-18>
- Tejada, E., Garay, U., Romero, A., & Bilbao, N. (2021). El bullying desde el punto de vista del acosador: Análisis y procedimiento. *Revista de Investigación Educativa*, 39(2), 373-390. <https://doi.org/10.6018/RIE.422671>
- Ubieto-Pardo, J. R. (2021). Los terceros en el bullying: ¿testigos o cómplices? *Revista de la Asociación Española de Neuropsiquiatría*, 41(139), 267-278. <https://www.revistaaen.es/index.php/aen>
- Vasilachis de Gialdino, I. (Ed.) (2006). *Estrategias de investigación cualitativa*. Gedisa.

ANEXO 1

Guía para la realización de un juicio de expertos

“Cuestionario de Acoso y Ciberacoso Escolar (CACE)”

Evaluación de Expertos

Estimado juez: Usted ha sido seleccionado para evaluar el instrumento denominado “Cuestionario de Acoso y Ciberacoso Escolar (CACE)”. Su aporte es de gran importancia para que los resultados que se obtengan de la aplicación, sean pertinentes con las variables que se desean medir. Gracias por su colaboración.

Datos del Experto:

Nombre y apellido:

Formación Académica:

Área de Experticia:

Dependencia institucional:

El objetivo del test

i. Objetivo general:

- 1) Identificar la existencia de situaciones de acoso escolar.

i. Objetivos desagregados:

- 2) Identificar perfiles de los participantes (víctima, acosador, testigo).
- 3) Identificar tipos de acoso (indirecto, directo, social, verbal, cibernético).

Objetivo de la evaluación

- i. Identificar coherencia entre indicadores presentados en los ítems y las variables que pretenden medir.
- ii. Corroborar la adecuación del lenguaje utilizado en cada ítem en relación a la población a la que se dirige el instrumento.

Características de la población de aplicación

El instrumento está diseñado para estudiantes de nivel secundario común desde primero a sexto año. Su aplicación podría extenderse a estudiantes de sexto y séptimo grado de primaria. El constructo está diseñado para el rango de edad que va desde los 11 años a los 18 años.

1. Planilla de variables, dimensiones e indicadores del instrumento

Variable	Dimensión	Indicador
1. Violencia física indirecta	Acoso indirecto	Esconder
		Robar
		Romper
2. Violencia física directa	Opresión	Impedir que vaya al baño, al patio, al curso, etc.
		Impedir que pase por el pasillo u otro lugar.
		Encerrar en el baño, armario, habitación, etc.
		Otro:
	Agresión Física	Pegar (piña, cachetadas, coschacho, etc.)
		Empujar
		Dar patadas
		Escupir
		Poner zancadillas
		Amenazar con armas: revólver, cuchillo, objeto punzante, cutter, etc.
		Encerrar en el baño u otro espacio
		No dejar pasar
		Tirón de pelos
		Lanzar objetos
		Capacho
Otros.....		
3. Violencia verbal gestual directa	Humillación	Burlas
		Sarcasmo
		Ridiculizar
		Poner nombres despectivos (Sobrenombres)
		Dedicar canciones
		Imitación
	Insultar	Insultar
	Amenazar	Amenazar
Extorsión	Extorsionar	

Variable	Dimensión	Indicador
4. Violencia social relacional indirecta	Exclusión	Rechazo
		Discriminación: posibles causas de la discriminación: raza, religión, nerd, etc., pero no medimos la causa sino hecho fáctico de discriminación. Existe discriminación sí/no.
	Injuriar	Calumniar
		Mentir
		Inventar historias falsas
		Chismorrear, murmurar
Otro...		
5. Violencia Ciberacoso a. Violencia verbal gestual indirecta b. Violencia social relacional indirecta	a. Humillación	Burlas
		Sarcasmo
		Ridiculizar (fotos – videos – memes)
		Poner nombres despectivos (Sobrenombres)
		Robar acceso a cuentas
		Imitación
	a. Insultar	Insultar - memes
	a. Amenazar	Amenazar (fotos – videos)
	a. Extorsión	Extorsionar (fotos – videos)
	b. Exclusión	Rechazo
		Discriminación: posibles causas de la discriminación: raza, religión, nerd, etc., pero no medimos la causa sino hecho fáctico de discriminación. Existe discriminación sí/no.
	b. Injuriar	Calumniar
		Mentir
		Inventar historias falsas
Chismorrear		
Otro...		

2. Criterios para evaluar el Instrumento

Referencias de calificación:

- 1) No cumple con el criterio
- 2) Bajo Nivel
- 3) Moderado nivel
- 4) Alto nivel

De acuerdo con los siguientes indicadores califique cada uno de los ítems según corresponda.

Categoría	Referencia	Valor
SUFICIENCIA (Suf.): Los ítems que pertenecen a una misma dimensión o variable bastan para obtener la medición de ésta.	Los ítems no son suficientes para medir la dimensión.	1
	Los ítems miden algún aspecto de la dimensión, pero no corresponden con la dimensión total.	2
	Se deben incrementar algunos ítems para poder evaluar la dimensión completamente.	3
	Los ítems son suficientes.	4

Categoría	Referencia	Valor
CLARIDAD (Clar.): El ítem se comprende fácilmente, es decir, su sintaxis y semántica son adecuadas.	El ítem no es claro.	1
	El ítem requiere bastantes modificaciones en el uso de las palabras de acuerdo con su significado o el orden de las mismas.	2
	Se requiere una modificación muy específica de algunos de los términos del ítem.	3
	El ítem es claro, tiene semántica y sintaxis adecuada.	4
COHERENCIA (Coh.): El ítem tiene relación lógica con la dimensión o indicador que está midiendo.	El ítem no tiene relación lógica con la dimensión.	1
	El ítem tiene una relación tangencial con la dimensión.	2
	El ítem tiene una relación simple con la dimensión que está midiendo.	3
	El ítem se encuentra completamente relacionado con la dimensión que está midiendo.	4
RELEVANCIA (Rel.): El ítem es esencial o importante, es decir debe ser incluido.	El ítem puede ser eliminado sin que se vea afectada la medición de la dimensión.	1
	El ítem tiene alguna relevancia, pero otro ítem puede estar incluyendo lo que mide éste.	2
	El ítem tiene relevancia y puede ser incluido.	3
	El ítem es muy relevante (esencial) y debe ser incluido.	4

Nota. Planilla de juicio de expertos. Tomado de Validez de contenido y juicio de expertos: una aproximación a su utilización (p. 35) por Escobar-Pérez y Cuervo-Martínez (2008).

3. Planilla de evaluación de ítems

Dimensión	Ítem	Puntaje				Observaciones
		Suf.	Clar.	Coh.	Rel.	
Acoso Indirecto	1					
	2					
	3					
	4					
	5					
	6					
Opresión	7					
	8					
Agresión Física	9					
	10					
	11					
	12					
	13					
	14					
Humillación	15					
	16					
	17					

Dimensión	Ítem	Puntaje				Observaciones
		Suf.	Clar.	Coh.	Rel.	
Insulto	18					
	19					
	20					
Amenaza	21					
	22					
Extorsión	23					
	24					
Exclusión	25					
	26					
	27					
	28					
	29					
	30					
	31					
Injuriar	32					
	33					
	34					
(Ciberacoso) a. Humillación	35					
	36					
	37					
	38					
	39					
	40					
	41					
	42					
a. Insulto	43					
	44					
	45					
a. Amenaza	46					
a. Extorsión	47					
b. Exclusión	48					
b. Injuriar	49					
	50					
	51					
	52					
	53					
	54					