



Revisión de ítems: Validación del Cuestionario para la Evaluación del Estrés (tercera versión), de la Batería de Instrumentos para la Evaluación de Factores de Riesgo Psicosocial

Juan Esteban Gallego Duque

Trabajo de grado presentado para optar al título de Psicólogo

Tutor

Carlos Alberto Marín Monsalve, Esp. en Seguridad y Salud en el Trabajo

Universidad de Antioquia

Facultad de Ciencias Sociales y Humanas

Pregrado en Psicología

Medellín, Antioquia, Colombia

2022

Cita	(Gallego Duque, 2021)
Referencia	Gallego Duque, J. E. (2021). <i>Revisión de ítems: Validación del Cuestionario para la Evaluación del Estrés (tercera versión), de la Batería de Instrumentos para la Evaluación de Factores de Riesgo Psicosocial</i> [Trabajo de grado profesional]. Universidad de Antioquia, Medellín, Colombia.
Estilo APA 7 (2020)	



Créditos

Asesoras: Katherine Vallejo y Nancy Carmona, Psicólogas Esp. en Psicometría.

Jueces expertos: Diana Alexandra Mora Pedroza, Diego Castrillón Moreno, Esneda Bermúdez Gutiérrez, Jonathan Andrés Hernández Calle, Luana Betsy Polo Cortés, Luz Mayra Sanguino Cotte, María Alejandra Gómez Vélez, Miryam León Dimas, Nancy Alvarado Avila y Rosi Marcela Pérez Vera.



CRAI María Teresa Uribe (Facultad de Ciencias Sociales y Humanas)

Repositorio Institucional: <http://bibliotecadigital.udea.edu.co>

Universidad de Antioquia - www.udea.edu.co

Rector: John Jairo Arboleda Céspedes.

Decano: John Mario Muñoz.

Jefe de departamento: Alberto Ferrer Botero.

El contenido de esta obra corresponde al derecho de expresión de los autores y no compromete el pensamiento institucional de la Universidad de Antioquia ni desata su responsabilidad frente a terceros. Los autores asumen la responsabilidad por los derechos de autor y conexos.

Resumen

El Ministerio de Protección Social a partir de la resolución 2648 de 2008, señala la importancia del cuidado de la salud física y mental en los entornos laborales, momento a partir del cual se desarrolla la batería de Instrumentos para la Evaluación de Factores de Riesgo Psicosocial que incluye, a partir del 2010, la tercera versión del cuestionario para la Evaluación del Estrés. Por inconsistencias evidenciadas en el cuestionario, se determinó como objetivo central, la revisión y validación de sus ítems, con el propósito de evaluar los aspectos susceptibles de mejora en su formulación.

Lo primero en realizarse fue una profunda revisión teórica bajo el paradigma interpretativo, y lo segundo un juicio de expertos con el fin de validar la suficiencia, claridad, coherencia y relevancia de las dimensiones que constituyen el cuestionario. Los datos obtenidos mediante los jueces fueron usados para analizar tres aspectos: 1) Índices de Validez de Contenido (IVC); 2) consistencia interna por Alfa de Cronbach; y 3) consistencia externa por W de Kendall.

Los IVC del cuestionario se encuentran entre considerables y casi perfectos, pues ninguno presenta un porcentaje inferior al 61%, por otra parte, la consistencia interna elimina directamente 5 ítems al encontrarse en un rango inferior al 0.05, y la consistencia externa invita a modificar o eliminar la totalidad de los ítems del cuestionario al no haber ninguno superior al 0,6. Finalmente es de señalar que, aunque un instrumento obtenga una alta evaluación durante su validación, debe permanecer en continua revisión y mejoramiento.

Palabras clave: revisión de ítems, validación, factores de riesgo psicosocial, cuestionario de estrés, juicio de expertos, alfa de cronbach, w de kendall, spss

Abstract

The Ministry of Social Protection, with the resolution N° 2648 (2008), points out the importance of physical and mental health care in work environments, at which point the battery of Instruments for the evaluation of psychosocial risk factors is developed, includes, since 2010, the third version of the quiz for the Stress Assessment. Due to inconsistencies evidenced in the quiz, the main objective was to review and validate those items, with the purpose of evaluating the susceptible aspects to improvements in its formulation.

The first thing we did was a deep theoretical review under the Interpretive Paradigm, and the second one was an expert judgment in order to validate the sufficiency, clarity, coherence and relevance of the dimensions that made up the quiz. The data obtained by the judges, were used to analyze three aspects: 1) Content Validity Indexes (CVI); 2) internal consistency by Cronbach's Alpha; and 3) external consistency by Kendall's W.

The CVI of the questionnaire are between considerable and almost perfect indexes, since none presents a percentage lower than 61%, secondly the internal consistency directly eliminates 5 items, it is range lower than 0.05, and the external consistency invites us to modify or remove all the items from the quiz as there were none greater than 0,6. Finally, it should be noted that, although an instrument obtains a high evaluation during its validation, it must have a continuous review and improvement.

Keywords: review of ítems, validation, psychosocial risk factors, stress questionnaire, expert judgment, cronbach's alpha, w for kendall, spss

Introducción

Actualmente, señalan Díaz y Rentería (2017), empresas e instituciones están inmersos en continuos y acelerados cambios que se enmarcan en los procesos de globalización e interdependencia económica que transforma el mundo laboral. Aunado a ello, predominan las culturas organizacionales que carece de efectividad en el manejo de los recursos humanos, con políticas y prácticas que generan riesgos y trastornos psicosociales que inciden en la eficacia del trabajo, y que además, impactan en la salud física y emocional de los trabajadores, produciendo ansiedad y otros factores desencadenantes del estrés (Salamanca Velandia, Pérez Torres, Infante Alvarado, & Olarte Ardila, 2019).

El Ministerio de la Protección Social, expide la Resolución 2646 (2008) que dicta de manera técnica y puntual, que el estrés laboral es la “respuesta de un trabajador tanto a nivel fisiológico, psicológico como conductual, en su intento de adaptarse a las demandas resultantes de la interacción de sus condiciones individuales, intralaborales y extralaborales”, señalando además, que los factores psicosociales deben ser evaluados objetiva y subjetivamente, utilizando instrumentos que hayan sido validados en el país (Ministerio de La Protección Social, 2008).

Por este motivo indican Charria et al. (2011), se construye la Batería de Instrumentos para la Evaluación de Factores de Riesgo Psicosocial (BIEFRP), lo cual implicó la revisión de una gran cantidad de literatura científica y técnica sobre factores psicosociales, y sobre los métodos e instrumentos para su evaluación, planteando el modelo conceptual y las metodologías de medición de su estructura. La Batería incluyó el cuestionario para la Evaluación del Estrés en su tercera versión (Gómez, Segura Camacho, Castrillón, & Perilla, 2016).

La versión inicial de este cuestionario (1996), se construyó y validó con una muestra de 172 personas con edades promedio de 35 años; la segunda versión (2005) implicó importantes cambios en el sistema calificativo de niveles de riesgo para la salud (al igual que en la baremación) gracias a un estudio de seguimiento a cohortes de trabajadores; la tercera y última versión, se validó con 2 199 trabajadores de diversas actividades económicas y ocupaciones, y los baremos se construyeron para cuatro niveles ocupacionales: Directivo, profesional, auxiliar, y operativo (Ministerio de La Protección Social, 2010).

En este cuestionario, el estrés es definido entorno al modelo denominado Transaccional de Lazarus y Folkman (1984), haciendo referencia a las interrelaciones que se generan entre la persona y su contexto, afirmando que el estrés se produce cuando la persona valora lo que sucede como algo que supera los recursos con los que cuenta, poniendo en peligro su bienestar personal con base en la evaluación cognitiva que realiza, contemplando también el factor emocional que conlleva a esta situación (de Vera Estévez, 2004).

Velázquez et al. (2008) de manera acorde, definen estrés laboral como el conjunto de fenómenos ocurridos en el organismo del trabajador como consecuencia de la aparición de elementos nocivos, que se derivan directamente de su trabajo. La reacción que un trabajador puede presentar como consecuencia de las exigencias y presiones que no son adecuadas para los conocimientos que posee, y que ponen a prueba su capacidad para hacerles frente, propician el estrés laboral, como resultado de la dificultad o incapacidad para afrontar estas situaciones (Gutiérrez Strauss & Vilorio Doria, 2014).

El Cuestionario para la Evaluación del Estrés Versión Tres (CEE-V3) se divide en 31 síntomas agrupados, presentando un ítem por síntoma, por lo cual debe administrarse en su totalidad para que sea posible obtener resultados válidos. Los síntomas que revelan la presencia de estrés han sido distribuidos en cuatro dimensiones: a) fisiológicos (8 ítems), b) comportamiento social (4 ítems), c) intelectuales y laborales (10 ítems), y d) psicoemocionales (9 ítems) (Ministerio de La Protección Social, 2010).

Según Muñiz (2010), la distribución de los ítems en el cuestionario plantea un inconveniente, ya que todas las dimensiones que representan la variable deben estar igualmente representados, evitando la infra o sobrerrepresentación, con aproximadamente el mismo número de ítems. Se aceptan como excepción, casos en los cuales haya dimensiones consideradas más relevantes que otras dentro de la variable, por lo cual sea necesario un mayor número de ítems para contar así con mayor representatividad, justificada por el modelo teórico subyacente y la evidencia empírica acumulada (Muñiz & Fonseca Pedrero, Diez pasos para la construcción de un test, 2019).

Por este y otros aspectos observables en el CEE-VE, se determinó la revisión y validación de sus ítems como objetivo central, con el propósito de evaluar los aspectos susceptibles de mejora en la formulación de estos, teniendo en cuenta la identificación del modelo teórico usado

para su construcción, y de la bibliografía existente alrededor del tema, pero centrados únicamente en los ítems, y sin pretender (en primera instancia) el desarrollo de un instrumento nuevo.

Para el 2010, Prieto y delgado definen la validez de contenido como “la justificación de que los ítems para medir el criterio son una muestra representativa del contenido a evaluar”. Sin embargo y más recientemente (2017), Koller et al. actualizan dicha definición al indicar que, “la validez del contenido es determinada por la validez y representatividad en la definición del constructo, representatividad del grupo de ítems, aspectos gramaticales de los ítems y claridad de las instrucciones (Juárez Hernández & Tobón, 2018)”.

Es así como los ítems son apenas uno de los múltiples elementos que constituyen un instrumento evaluativo según Murillo (2006), por ende, el enfoque fue puesto en la validez de contenido y la validez de constructo, siendo la primera evidencia a favor de la validez general de una prueba. Para la revisión se realizó un análisis de contenido, que determinó la representatividad y redacción de cada uno de los ítems mediante un juicio de expertos, que como resultado final sugerirá modificaciones o ajustes de constructo (Bolado, Ibáñez, & Lantarón, 1998).

La construcción de los ítems es crucial durante el proceso de elaboración de cuestionarios, siendo estos la materia prima o columnas a partir de las cuales se conforma un instrumento de evaluación, indica Morales (2010). Una construcción deficiente de los ítems incidirá en las propiedades métricas finales del instrumento de medida, y en la validez de las inferencias que se hagan a partir de las puntuaciones (Muñiz & Fonseca Pedrero, Diez pasos para la construcción de un test, 2019).

La validez de los tests se confirma con estudios experimentales, aunque se puede hablar de una validación previa de tipo conceptual o cualitativo como parte del proceso de validación, y para asegurar que en principio los ítems son claros y el instrumento mide lo deseado, según Morales (2011). Por ello, el primer control respecto a la validez está en la redacción misma de los ítems (García Sánchez, Aguilera Terrats, & Castillo Rosas, 2011).

Por otra parte, la confiabilidad y la validez del cuestionario fueron estimadas de formas diversas y complementaria en las tres versiones, y los ítems se diseñaron bajo el modelo de escala de respuestas tipo Likert (siempre; casi siempre; a veces; nunca), facilitando con ello que las

personas seleccionen una única opción de respuesta que, dentro de lo esperado, será la que mejor refleje o se ajuste, a la ocurrencia reciente de algunos de los síntomas evaluados (Matas, 2018).

Cuando se establece la necesidad de medir, como en el caso puntual del CEE-V3, se requiere que el instrumento de medición sea objetivo y permita realizar comparaciones, obteniendo valores numéricos (directa o indirectamente) con los cuales se puedan hacer ordenamientos que faciliten, dentro de la complejidad que esto involucra, la valoración de los rasgos medidos en las personas, en este caso, rasgos de estrés en el ámbito laboral (Tristán López & Molgado Ramos, 2007).

Las escalas tipo Likert hacen parte de las pruebas de ejecución típica (o de ejecución de rasgos), que son aquellas que miden aspectos no escalables, y cuyas preguntas no tienen respuestas correctas o erróneas, sino que tratan aspectos de opinión, de preferencia, o similitud (González Romá, 2001). Para Meneses (2013), este tipo de escalas se ha convertido en el más popular debido a su facilidad de aplicación y a su adaptación a cualquier ámbito de evaluación, imponiéndose ante otros modelos mejor fundamentados científicamente, pero de aplicación menos sencilla.

Metodología

En este proceso se llevó a cabo, en primera instancia, una profunda revisión teórica alrededor del diseño y la medición de ítems, para así tener claridad y precisión respecto a la validez y confiabilidad del cuestionario para la Evaluación del Estrés en su tercera versión (CEE-V3), como lo señalan Urrutia et al. (2014). En segunda instancia, se llevó a cabo un juicio de expertos, en búsqueda de argumentos que demuestren la suficiencia, claridad, coherencia y relevancia de los ítems existentes (Dorantes Nova, Hernández Mosqueda, & Tobón Tobón, 2016).

La revisión teórica se basó en el paradigma interpretativo, donde la tarea del investigador es estudiar el proceso interpretativo que los expertos hacen de los ítems diseñados para medir el fenómeno, investigando el modo en que se asignan significados, conceptos, y teorías sobre dichos ítems, permitiendo medir de la manera más transparente y precisa el fenómeno en cuestión (Galeano Marín, Estrategias de investigación social cualitativa: El giro de la mirada, 2009). Este enfoque implica, señala Krause (1995), interpretar desde la perspectiva de los autores, enfatizando el proceso de comprensión en el investigador.

Como estrategia metodológica, fue usada la investigación documental, que implica un esfuerzo por mostrar desde la heterogeneidad de los materiales documentales que circulan en el medio sobre el tema, una trama de relaciones y conexiones sobre los aspectos presentes en los mismos, permitiendo de esta manera establecer un orden jerárquico, señalando vacíos y necesidades de articulación, haciéndolas viables y accesibles (Galeano Marín, Estrategias de Investigación Social Cualitativa: El giro de la mirada, 2009).

Se hizo también uso en esta primera parte del proceso, de un diseño cualitativo, en el cual se prescinde de datos cuantificables, basándose específicamente en la descripción de cualidades mediante conceptos y relaciones entre ellos, lo que es definido por Hernández et al. (2014), como aquel diseño que puede desarrollar preguntas o hipótesis en varios momentos de la investigación, es decir, antes, durante, o después de la recolección y análisis de los datos.

El juicio de expertos, usado en segunda instancia, es una opinión informada de personas con trayectoria en el tema y reconocidos por otros profesionales como expertos cualificados en

la materia (por su formación y/o experiencia), que pueden brindar información, evidencia, juicios y valoraciones precisas (Galicia Alarcón, Balderrama Trápaga, & Edel Navarro, 2017). Robles y Rojas (2015), plantean que el juicio de expertos en muchas disciplinas como la Psicología, es una parte importante de la información cuando las observaciones experimentales son limitadas.

El número de expertos a usar en un juicio depende del nivel de experticia y de la diversidad del conocimiento, sin embargo, mientras Gable y Wolf (1993), Grant y Davis (1997), y Lynn (1986) sugieren un rango de dos hasta veinte expertos; Hyrkäs et al. (2003) manifiestan que diez brindarían una estimación confiable de la validez de contenido de un instrumento. Al final todos coinciden en que, si un 80 % de los expertos han estado de acuerdo con la validez de un ítem, éste puede incorporarse al instrumento (Escobar Pérez & Cuervo Martínez, Validez de contenido y juicio de expertos: Una aproximación a su utilización, 2008).

En línea con Hyrkäs et al. (2003), se seleccionó un grupo de diez expertos con experiencia en los campos de la Psicometría (P), la Seguridad y Salud en el Trabajo (SST), y los Riesgos Psicosociales (RP) durante un periodo de tiempo mínimo de dos años. Ninguno de ellos, al momento de llevar a cabo el ejercicio de forma individual, tiene relación directa con el diseño original del cuestionario y/o la creación de la batería misma, señalando también que se respetaron los demás protocolos éticos y de confidencialidad establecidos para situaciones de esta índole (Ministro de Salud y Protección Social, 2016).

Tabla 1: *Perfil de los jueces expertos*

JUEZ	SEXO	CAMPO OCUPACIONAL	NIVEL PROFESIONAL
1	M	P	Doctorado
2	F	SST	Especialización
3	F	SST	Especialización
4	F	SST	Especialización
5	F	P	Maestría
6	F	RP	Especialización
7	M	P	Maestría
8	F	SST	Especialización
9	F	RP	Maestría
10	F	RP	Especialización

El instrumento brindado a los expertos para la validación cubre cuatro categorías o criterios para cada uno de los 31 ítems del CEE-V3 en sus cuatro dimensiones, fue elaborado virtualmente en la plataforma JotForm (JotForm Inc., 2021), y se les compartió vía correo electrónico previa autorización de cada uno de ellos. Se utilizó la escala Likert para dicho ejercicio (Abad, Garrido, Olea, & Ponsoda, 2006).

Se brindó también a los expertos, un apartado para su apreciación cualitativa al final de cada una de las dimensiones del cuestionario, propiciando así el enriquecimiento descriptivo en favor de los ítems existentes en él, apuntando a múltiples aspectos relevantes como la complejidad del ítem alrededor de la población a la que va dirigido el cuestionario, y el abordaje de la dimensión a la cual pertenece cada ítem (Pereira Pérez, Los Diseños de Método Mixto en La Investigación en Educación: Una experiencia concreta, 2011). Además de cualquier otro tipo de sugerencia u observación, la cual pueda ser importante a lo largo del proceso, o en futuros proyectos relacionados a este.

Tabla 2: *Método de calificación*

CATEGORÍA	CALIFICACIÓN	INDICADOR
Suficiencia: Los ítems que pertenecen a una misma dimensión bastan para obtener la medición de ésta.	1. No cumple el criterio	Los ítems no son suficientes para medir la dimensión.
	2. Bajo nivel	Los ítems miden algún aspecto de la dimensión, pero no corresponden con la dimensión total.
	3. Moderado nivel	Se deben incrementar algunos ítems para poder evaluar la dimensión completamente.
	4. Alto nivel	Los ítems son suficientes.
Claridad: El ítem se comprende fácilmente, es decir, su sintáctica y semántica son adecuadas.	1. No cumple el criterio	El ítem no es claro.
	2. Bajo nivel	El ítem requiere bastantes modificaciones o una modificación muy grande en el uso de las palabras de acuerdo con su significado o por la ordenación de las mismas.
	3. Moderado nivel	Se requiere una modificación muy específica de algunos de los términos del ítem.
	4. Alto nivel	El ítem es claro, tiene semántica y sintaxis adecuada.
Coherencia: El ítem tiene relación lógica con la dimensión o indicador que está midiendo.	1. No cumple el criterio	El ítem no tiene relación lógica con la dimensión.
	2. Bajo nivel	El ítem tiene una relación tangencial con la dimensión.
	3. Moderado nivel	El ítem tiene una relación moderada con la dimensión que está midiendo.
	4. Alto nivel	El ítem se encuentra completamente relacionado con la dimensión que está midiendo.
Relevancia: El ítem es esencial o importante, es decir, debe ser incluido.	1. No cumple el criterio	El ítem puede ser eliminado sin que se vea afectada la medición de la dimensión.
	2. Bajo nivel	El ítem tiene alguna relevancia, pero otro ítem puede estar incluyendo lo que mide éste.
	3. Moderado nivel	El ítem es relativamente importante.
	4. Alto nivel	El ítem es muy relevante y debe ser incluido.

Para la validación y mediante los datos obtenidos, se revisaron tres aspectos:

1. Índices de Validez de Contenido (IVC) de los ítems. Se considera aceptable cuando la favorabilidad del ítem supera el 80 % entre los expertos, aquellos que no alcanzan este porcentaje se recomienda sean eliminados y remplazados en el instrumento o,

simplemente modificados y sometidos a un nuevo proceso de validación (Hernández González, Ortega Andeane, & Reidl Martínez, 2012).

2. Consistencia interna con coeficientes de fiabilidad Alfa (α) de Cronbach. Tienen como objetivo seleccionar los ítems que mejor contribuyen a la fiabilidad de toda la escala, entendiendo que un coeficiente de fiabilidad alto apoya, pero no prueba, que todos los ítems miden lo mismo o expresen bien el mismo rasgo, ya que una fiabilidad alta expresa simplemente que los ítems están relacionados entre sí (Gómez Benito & Hidalgo Montesinos, 2003).
3. Consistencia externa con coeficiente de concordancia de Kendall (W). Se obtiene a partir del análisis de cada ítem, empleando los criterios de suficiencia, claridad, coherencia y relevancia, determinando con ello el grado de acuerdo entre los jueces o expertos, en un rango de 0 a 1 (Verlanga Silvente & Rubio Hurtado, 2012).

Los datos obtenidos se tabularon en el software contable Excel (Microsoft, 2016), con el cual también se obtuvieron los promedios para el primer aspecto a revisar, y se realizaron parte de las tablas explicativas y descriptivas de la validación. Las tablas restantes y los dos aspectos más a revisar (Alfa de Cronbach y W de Kendall), fueron analizados con el software estadístico SPSS (IBM, 2013).

Resultados

El análisis de resultados supone, en primera instancia, calcular la concordancia promedio entre los jueces para cada uno de los ítems en cada categoría evaluada, para así obtener el IVC que permite dar pie a las primeras impresiones respecto al cuestionario.

Tabla 3: *Porcentaje de IVC*

DIMENSIONES	ÍTEMS	SUFICIENCIA	CLARIDAD	COHERENCIA	RELEVANCIA	TOTAL (IVC)
Fisiológicos	1	85%	92,5%	90%	97,5%	91,3%
	2	90%	90%	90%	100%	93%
	3	67,5%	77,5%	85%	90%	80%
	4	87,5%	92,5%	97,5%	100%	94%
	5	87,5%	95%	95%	97,5%	93,8%
	6	85%	80%	90%	97,5%	88,1%
	7	85%	85%	90%	90%	88%
	8	77,5%	80%	85%	80%	81%
Comportamiento Social	9	82,5%	75%	90%	92,5%	85,0%
	10	72,5%	70%	75%	80%	74%
	11	70%	62,5%	77,5%	80%	73%
	12	77,5%	77,5%	80%	82,5%	79,4%
Intelectuales y Laborales	13	87,5%	85%	87,5%	95%	89%
	14	90%	92,5%	95%	92,5%	92,5%
	15	77,5%	75%	85%	90%	82%
	16	87,5%	87,5%	92,5%	95%	91%
	17	77,5%	87,5%	80%	87,5%	83,1%
	18	85%	82,5%	92,5%	95%	89%
	19	85%	90%	95%	97,5%	91,9%
	20	85%	87,5%	92,5%	95%	90%
	21	87,5%	87,5%	95%	97,5%	91,9%
	22	85%	85%	90%	92,5%	88,1%
Psicoemocionales	23	90%	92,5%	95%	97,5%	93,8%
	24	82,5%	77,5%	90%	92,5%	85,6%
	25	90%	95%	97,5%	97,5%	95,0%
	26	75%	70%	87,5%	92,5%	81,3%
	27	90%	92,5%	95%	97,5%	93,8%
	28	67,5%	55%	87,5%	90%	75%
	29	82,5%	87,5%	90%	95%	89%
	30	80%	85%	87,5%	90%	86%
	31	90%	95%	95%	95%	94%

Entendemos como “considerables” (rango entre 61% - 80%) los ítems 11, 10, 28, 12 y 3, que en ese mismo orden presentan los IVC más bajos de toda la prueba, resaltando el hecho de que tres de ellos hacen parte de la dimensión Comportamiento Social, la cual está conformada por tan solo cuatro ítems, siendo sin duda la dimensión más baja en general con un 77,8% (la más alta es Intelectuales y Laborales con el 89%). Los demás ítems al hallarse sobre el 81% de IVC, son “casi perfectos” en cuanto a términos generales se refiere, siendo el criterio de Suficiencia el más bajo a nivel general con un 83% de IVC, y el de Relevancia el mayor con un 92,7% (Bernal García, Salamanca Jiménez, Perez Gutiérrez, & Quemba Mesa, 2018).

La consistencia interna por Alfa (α) de Cronbach, nos señala el grado en que la totalidad de los ítems covarían entre sí, esto indica hasta qué punto medidas parciales obtenidas con los múltiples ítems, son representativas del universo posible de ítems que pueden medir ese constructo. La fiabilidad de un instrumento establece la precisión con la cual este lleva a cabo sus mediciones, siendo un concepto asociado con el error de medida, ya que a mayor fiabilidad menor es el error en la medición (Hernández González, Ortega Andeane, & Reidl Martínez, 2012).

Tabla 4: Alfa (α) de Cronbach Dimensiones/Categorías

DIMENSIÓN	α SUFICIENCIA	α CLARIDAD	α COHERENCIA	α RELEVANCIA	α TOTAL
Fisiológicos	0,270	0,009	0,24	0,023	0,001
Comportamiento Social	0,297	0,298	0,280	0,281	0,008
Intelectuales y Laborales	0,535	0,199	0,186	0,717	0,000
Psicoemocionales	0,035	0,000	0,322	0,833	0,000
α TOTAL	0,156	0,000	0,117	0,130	

Los resultados de α de Cronbach son deficientes en la totalidad de dimensiones que hacen parte del CEE-V3, al igual que en todos los criterios o categorías evaluadas. Sin embargo para Navarro et al. (2020), es un resultado poco diciente en este punto de la validación, pues lo ideal para estos casos y para dicha estadística, es contar con un instrumento alternativo o una doble evaluación del instrumento existente mediante un test retest, o a través de dos o más grupos de jueces trabajando de forma separada, por ejemplo.

Es así como tanto la consistencia interna como la consistencia externa, se llevaron a cabo con el análisis estadístico de la información obtenida, en la cual el 1 representa total concordancia

y el 0 un pleno desacuerdo, por consiguiente, entre mayor sea el valor del Alfa (α) de Cronbach o el W de Kendall, más fuerte es la validez y confiabilidad del instrumento respecto a los fines para los que fue diseñado (Vásquez Gonzáles, Jiménez Macías, & Juárez Hernández, 2020).

Puntualizando cada ítem, es evidente que las categorías cuentan con al menos un ítem de Hipótesis Nula (H_0), implicando esto su eliminación inmediata al encontrarse en un rango inferior al 0.05 como señalan Escobar y Cuervo (2008). Pero, a pesar de que la mayoría de ítems aceptan la Hipótesis Alternativa (H_1) que les permite ser aceptados dentro del instrumento, solo unos pocos se encuentran cercanos a 1 en un rango superior al 0.7, lo cual indica que a excepción del ítem 8, todos los demás deben ser modificados con base a su consistencia interna, siendo preocupante en términos porcentuales, que el 25% de la dimensión de Comportamiento Social no es válida, y el 75% restante debe modificarse, aunque se precisa ser cambiado como lo indican Zamora et al. (2020).

Tabla 5: Alfa (α) de Cronbach ítems

DIMENSIÓN	ÍTEM	α	ρ	HIPÓTESIS ACEPTADA
Fisiológicos	1	0,264	$\rho > 0,05$	H□
	2	0,147	$\rho > 0,05$	H□
	3	0,071	$\rho > 0,05$	H□
	4	0,072	$\rho > 0,05$	H□
	5	0,096	$\rho > 0,05$	H□
	6	0,035	$\rho < 0,05$	H□
	7	0,644	$\rho > 0,05$	H□
	8	0,807	$\rho > 0,05$	H□
Comportamiento Social	9	0,011	$\rho < 0,05$	H□
	10	0,241	$\rho > 0,05$	H□
	11	0,103	$\rho > 0,05$	H□
	12	0,300	$\rho > 0,05$	H□
Intelectuales y Laborales	13	0,212	$\rho > 0,05$	H□
	14	0,733	$\rho > 0,05$	H□
	15	0,042	$\rho < 0,05$	H□
	16	0,246	$\rho > 0,05$	H□
	17	0,392	$\rho > 0,05$	H□
	18	0,047	$\rho > 0,05$	H□
	19	0,120	$\rho > 0,05$	H□
	20	0,152	$\rho > 0,05$	H□
	21	0,090	$\rho > 0,05$	H□
	22	0,121	$\rho > 0,05$	H□
Psicoemocionales	23	0,547	$\rho > 0,05$	H□
	24	0,294	$\rho > 0,05$	H□
	25	0,603	$\rho > 0,05$	H□
	26	0,011	$\rho < 0,05$	H□
	27	0,204	$\rho > 0,05$	H□
	28	0,001	$\rho < 0,05$	H□
	29	0,139	$\rho > 0,05$	H□
	30	0,139	$\rho > 0,05$	H□
	31	0,392	$\rho > 0,05$	H□

El coeficiente de concordancia W de Kendall para consistencia externa, es una prueba no paramétrica, existiendo otras variables que no siguen las condiciones de parametricidad, las cuales aluden al uso de variables cuantitativas continuas como lo mencionan Berlanga y Rubio (2012). En las Ciencias Sociales estas pruebas son de uso habitual, permitiendo obtener rangos de variabilidad en la concordancia entre expertos, en cuyo caso y si la concordancia es baja, permite ajustar o eliminar el ítem hasta obtener el adecuado (Dorantes Nova, Hernández Mosqueda, & Tobón Tobón, 2016).

Tabla 6: *Concordancia W de Kendall*

Dimensiones	Ítem	W de Kendall	Concordancia Media de Categorías	Concordancia Media Total
FISIOLÓGICOS	1	0,132	0,170	0,185
	2	0,179		
	3	0,234		
	4	0,233		
	5	0,211		
	6	0,286		
	7	0,056		
	8	0,032		
COMPORTAMIENTO SOCIAL	9	0,374	0,211	
	10	0,140		
	11	0,206		
	12	0,122		
INTELECTUALES Y LABORALES	13	0,150	0,175	
	14	0,043		
	15	0,273		
	16	0,138		
	17	0,100		
	18	0,266		
	19	0,195		
	20	0,176		
	21	0,217		
	22	0,194		
PSICOEMOCIONALES	23	0,071	0,199	
	24	0,124		
	25	0,062		
	26	0,370		
	27	0,153		
	28	0,546		
	29	0,183		
	30	0,183		
	31	0,100		

Las puntuaciones bajas en cada ítem, afirman Morales y Vallejo (2011), corresponden a puntuaciones bajas en todos los ítems, motivo por el cual para poder afirmar con certeza que todos los ítems miden lo mismo, hacen falta controles conceptuales. Dichos controles demuestran baja rigurosidad en la construcción del CEE-V3, siendo esto evidente por formas y aspectos de redacción como lo son la doble negación y la inclusión de varios reactivos (y/o) en un mismo ítem, que desde el inicio desmejoraron la validez de constructo y de contenido del instrumento (Attorresi, Lozzia, Abal, Galibert, & Aguerri, 2009).

Los ítems 8 y 14 son los únicos que están debajo del rango aceptable de 0.05, motivo por el cual deberían ser descartados directamente, pudiéndose hacer lo mismo con el ítem 7 al

hallarse en un rango de valor crítico, y modificando de manera significativa los ítems 25 y 23 por sus bajos niveles. De acuerdo a los criterios de Tristán (2008), el valor 0,058 es el mínimo aceptable en un ítem, sin embargo Abdollahpour et al. (2011), defienden que solo los ítems superiores 0,8 deben conservarse en un instrumento, pudiéndose mejorar aquellos por encima del 0,7 y descartando sin duda aquellos que se encuentran por debajo de este valor (Zamora de Ortiz, Serrano Pastor, & Martínez Segura, 2020).

Tabla 7: Promedios

DIMENSIONES	ÍTEMS	SUFICIENCIA	CLARIDAD	COHERENCIA	RELEVANCIA	TOTAL
Fisiológicos	1	3,4	3,7	3,6	3,9	3,650
	2	3,6	3,6	3,6	4,0	3,700
	3	2,7	3,1	3,4	3,6	3,200
	4	3,5	3,7	3,9	4,0	3,775
	5	3,5	3,8	3,8	3,9	3,750
	6	3,4	3,2	3,6	3,9	3,525
	7	3,4	3,4	3,6	3,6	3,500
	8	3,1	3,2	3,4	3,2	3,225
Comportamiento Social	9	3,3	3,0	3,6	3,7	3,400
	10	2,9	2,8	3,0	3,2	2,975
	11	2,8	2,5	3,1	3,2	2,900
	12	3,1	3,1	3,2	3,3	3,175
Intelectuales y Laborales	13	3,5	3,4	3,5	3,8	3,550
	14	3,6	3,7	3,8	3,7	3,700
	15	3,1	3,0	3,4	3,6	3,275
	16	3,5	3,5	3,7	3,8	3,625
	17	3,1	3,5	3,2	3,5	3,325
	18	3,4	3,3	3,7	3,8	3,550
	19	3,4	3,6	3,8	3,9	3,675
	20	3,4	3,5	3,7	3,8	3,600
	21	3,5	3,5	3,8	3,9	3,675
	22	3,4	3,4	3,6	3,7	3,525
Psicoemocionales	23	3,6	3,7	3,8	3,9	3,750
	24	3,3	3,1	3,6	3,7	3,425
	25	3,6	3,8	3,9	3,9	3,800
	26	3,0	2,8	3,5	3,7	3,250
	27	3,6	3,7	3,8	3,9	3,750
	28	2,7	2,2	3,5	3,6	3,000
	29	3,3	3,5	3,6	3,8	3,550
	30	3,2	3,4	3,5	3,6	3,425
	31	3,6	3,8	3,8	3,8	3,750

Un análisis cuantitativo de las evaluaciones realizadas por los expertos para cada ítem logra determinar el grado de fiabilidad en cada uno de ellos y, a su vez, en cada una de las

dimensiones que conforman el cuestionario (Dorantes Nova, Hernández Mosqueda, & Tobón Tobón, 2016). A partir de las valoraciones en un rango de 1 a 4, y en función de cada criterio o categoría revisada dentro de la estructura, se evidencian altos puntajes en la mayoría, con una media de 3.48, una mediana de 3.55, y una moda de 3.75, siendo los ítems 11 y 25 el de mayor y menor promedio respectivamente, con una diferencia de casi 1 punto (0,9) entre ambos.

Discusión

Los factores psicosociales percibidos por los trabajadores como nocivos pueden incrementar al doble las posibilidades de un desgaste psíquico, según Carrión y Hernández (2018). De esta misma manera, el estrés puede ser leído en cuanto a sus posibilidades de afección en ámbitos humanos como el laboral, facilitando con ello su prevención mediante el uso de instrumentos de medición como baterías y cuestionarios, lo que obliga a los profesionales a mantener actualizadas y validadas este tipo de herramientas (Muñoz Rojas, Orellano, & Hernández Palma, 2018).

La validación de contenido de un instrumento articula una serie de elementos que requieren de una atención especial y diferencial. De esta manera, la caracterización y selección de jueces para un juicio de expertos, el empleo de un instrumento que permita la evaluación cuantitativa y cualitativa, y el análisis de los resultados de esta evaluación mediante coeficientes pertinentes para tal fin, representan fases de alta atención y la aplicación de métodos formales (Juárez Hernández & Tobón, 2018).

El IVC resultante del juicio de expertos de la validación del CEE-V3, muestra un nivel elevado en los ítems al estar por encima del 73% en términos generales, reflejando que, con simples ajustes en algunos de los ítems, sería posible consolidar el CEE-V3 como un instrumento válido, pero también preciso. Sin embargo, al mirar en detalle, es posible observar que la Suficiencia y la Claridad generales son cuestionables, reiterando como lo señalaron Attorresi et al. (2009) anteriormente, la importancia de una adecuada redacción de los ítems al inicio del proceso de construcción de un instrumento de este tipo.

Deben revisarse principalmente la sintaxis y la semántica de los criterios ya mencionados, y de los ítems 8, 10, 11, 12, 15 y 17 dentro de los dos criterios restantes que son Coherencia y Relevancia. Desde el principio la disparidad de ítems entre las cuatro dimensiones que constituyen el cuestionario llamó la atención, destacando que tres de los cuatro ítems que hacen parte de la dimensión de Comportamiento Social, se encuentran entre los más bajos de todo el instrumento, dando la razón a lo ya señalado por Muñiz (2010), quien recalca la importancia de contar con un número similar de ítems para cada dimensión.

Las ínfimas cifras por categorías del α de Cronbach, señalan deficiencias en el instrumento que, si bien son acordes en cierta medida con los otros dos aspectos de la validación, se limitan al grado de covarianza entre ítems, pues no se dispone de un instrumento o de una medición alterna del cuestionario que permita interpretar otros aspectos como la equivalencia o la estabilidad, para así arrojar cifras más precisas de la consistencia interna como lo indican Alfaro et al. (2020). El uso de herramientas complementarias, y de mediciones múltiples, es vital para dar certeza de la validez de un instrumento, sin embargo, siempre será propenso a mejoras de algún tipo (Meneses, y otros, 2013).

Por su parte, esta misma consistencia interna, señala que el cuestionario de *Evaluación del estrés* cuenta con únicamente un ítem fuerte (ítem 8) en este aspecto, siendo los demás ítems menos fiables, por lo cual requerirían de modificaciones, cambios o la eliminación inmediata, como es el caso de los ítems 6, 9, 15, 26 y 28, pudiéndose aprovechar estas modificaciones, para llevar a cabo una distribución más igualitaria entre dimensiones como aconsejan Muñiz y Fonseca (2019). Cabe señalar que la consistencia externa para el ítem 8 es aceptable más no elevada, con lo cual no existen ítems completamente sólidos dentro del instrumento.

La concordancia por W de Kendall para la consistencia externa puede entenderse como la de mayor precisión, ya que se parte de dos hipótesis, la nula y la alternativa. La primera se acepta para valores por debajo del 0.05, y señala la independencia de los rangos, motivo por el cual no concuerdan; la segunda indica que hay concordancia significativa entre los rangos y es válida para valores de 0.05 o más (Escobar Pérez & Cuervo Martínez, Validez de contenido y juicio de expertos: Una aproximación a su utilización, 2008).

Con base en lo anterior, cabe señalar que, aunque las puntuaciones por debajo del 0.05 se dan tan solo en 2 de los 31 ítems, solo 1 ítem sobrepasa los 0,5 en esta escala de 0 a 1, lo cual en términos generales refleja lo endeble del cuestionario, situación que no se evidenció con el análisis de las bases teóricas que sirvieron como pilares para la creación del instrumento. Esto invita a pensar en términos de Murillo (2006), quien recalca la importancia en la redacción de cada uno de los ítems que van a usarse en una prueba, dicho en otras palabras, es más un asunto de constructo que de contenido.

La revisión teoría que sustenta el CEE-V3 hace una importante distinción entre los tipos de estrés, y entre otros aspectos fuera del ámbito laboral que, de manera indirecta, repercuten en

los trabajadores con mayor o menor proporción e incidencia (Dahab, Rivadeneira, & Minici, 2010). De tal manera que la posibilidad siempre latente de crear un nuevo y mejorado instrumento, el cual pueda cubrir satisfactoriamente este tema dentro del área de los riesgos psicosociales, no tendría que partir necesariamente de cero.

Cabe mencionar la subjetividad que puede propiciarse entre los expertos que intervienen en un juicio, ya sea como proceso de creación o de validez de contenido como en este caso. Esta posible subjetividad propicia que los resultados se hallen sujetos a los sesgos que pueda haber entre ellos, para lo cual se considera correcta la identificación y operacionalización de los contenidos teóricos, compensando y propiciando así, el desarrollo exitoso del ejercicio, concluyen Zamora et al. (2020).

Conclusiones

Luego de los decretos y resoluciones dictaminadas por el gobierno hace unos años, entorno a los factores de riesgo psicosocial de origen laboral y a los altos niveles de estrés en el trabajo, el análisis de estos temas se ha acrecentado por parte de profesionales de distintas áreas, propiciando con ello el incremento en la rigurosidad alrededor de los instrumentos de evaluación para este tipo de asuntos. Así mismo, la oferta de herramientas de orígenes públicos y privados facilita el acceso que tienen a ellas las mipymes, quienes históricamente han sido las menos favorecidas ante este tipo de demandas legales.

El uso adecuado de las herramientas digitales facilita y acelera la puesta en marcha de ejercicios de validación o desarrollo de instrumentos de medición, propiciando la participación de profesionales desde cualquier parte del mundo, y generando mayor comodidad al momento de participar de evaluaciones como el juicio de expertos. De igual manera, los procesos de tabulación y análisis estadísticos se realizan cada vez más rápido, con mayor precisión, y a niveles altos de complejidad, gracias a los softwares que en el mercado se ofrecen para dichos fines.

No solo la cantidad de jueces es importante en la realización de un juicio de expertos, también sus años de experiencia o formación, y la relación directa que puedan tener con el aspecto a evaluar. Además, el análisis repetido y/o diferenciado de los instrumentos, y la redacción pertinente de los reactivos (ítems) según el contexto, también deben ser considerados fundamentales igual que el juicio de expertos al momento de crear o validar instrumentos de evaluación, de tal manera que los datos que estos provean al ser implementados sean altamente confiables desde el comienzo.

Los porcentajes arrojados para los IVC de los ítems, como resultado del juicio de expertos, no son equivalentes a los índices del Alfa de Cronbach o W de Kendall en cuanto al nivel de solidez de los ítems, por lo cual nunca debe hacerse uso de una única fórmula estadística al momento de llevar a cabo una validación. Por ejemplo, para esta revisión, el IVC señala que las dimensiones del cuestionario solo requieren del mejoramiento de algunos de sus ítems,

mientas que Alfa de Cronbach y W de Kendall, invitan tajantemente a descartarlos o eliminarlos todos, lo cual se traduce en la elaboración de un nuevo y distinto grupo de reactivos.

Las escalas Likert facilitan el diligenciamiento de encuestas, cuestionarios y demás instrumentos de medición, ya que agilizan las respuestas por parte de las personas al no tener que escribir, sino simplemente señalar la opción que mejor los representa o con la cual mejor se identifican. Sin embargo, errores de redacción, de selección de alternativas de respuesta, o de terminología empleada, pueden hacer que los datos obtenidos no sean acertados, hecho que también tiende a suceder, cuando las dimensiones de un instrumento no se encuentran repartidas de forma similar, provocando que la evaluación se polarice hacia uno o dos aspectos en vez de hacia un todo.

Alfa de Cronbach y W de Kendall de manera clásica avalan los ítems que se encuentran por encima del 0,05 en una escala de 0 a 1, con lo cual no requiere tan siquiera de un valor igual o superior al 50% dentro de la escala, generando ambigüedad, e invitando a replantearse la aprobación de ítems dentro de un margen superior al 0,7% como lo sostienen científicos e investigadores más modernos. Respecto al Alfa de Cronbach, es un estadístico que genera mayor confianza e información, cuando se usa con dos o más muestras diferentes en tiempo o espacio, motivo por el cual, en esta situación dichos datos son valorados con menor certeza respecto a los demás índices de esta validación.

Finalmente, Es fundamental mencionar que la validez de contenido es un componente importante de la validez de constructo, porque provee evidencia respecto al grado en el cual los elementos de un instrumento de evaluación son suficientes, claros, coherentes, y relevantes respecto al constructo objetivo, demostrando que la validez de contenido proporciona información sobre los datos obtenidos de un instrumento de evaluación, y las inferencias que se pueden extraer de esos datos. Se debe recordar que, aunque una prueba obtenga una muy buena evaluación de los jueces y un alto índice de concordancia, debe estar en continua revisión y mejoramiento.

Referencias

- Abad, F. J., Garrido, J., Olea, J., & Ponsoda, V. (2006). *Introducción a la psicometría: Teoría clásica de los tests y teoría de la respuesta al ítem*. Madrid, España: Universidad Autónoma de Madrid.
- Aguilera Vidal, H. N., Ramírez Chávez, E., & Torres Torres, R. I. (2019). Estrés laboral en guardias de seguridad. *Universidad y Sociedad, XI(5)*, 342-348.
- Aguirre Roldán, A., & Quijano Barriga, A. (2015). Síndrome por quemarse en el trabajo y variables familiares y laborales de los médicos generales de Bogotá. Una estrategia de calidad laboral. *Revista Colombiana de Psiquiatría, XLIV(4)*, 198-205.
- Alfaro Díaz, C., Canga Armayor, A., Gutiérrez Alemán, T., Carrión, M., & Esandi, N. (2020). Validación lingüística del Demand-Control-Support Questionnaire para profesionales de enfermería en España. *Anales del Sistema Sanitario de Navarra, XLIII(1)*, 17-33.
- Arce Julio, R., Rubio Buchard, K., Cuadro Vizcaino, H., Fonseca Angulo, R., León García, M., & Rodríguez Barraza, P. (2020). Relación entre el nivel de estrés laboral y los factores de riesgos psicosociales intralaborales en trabajadores de una fundación de niños con discapacidad ubicado en el departamento del Atlántico/Colombia. *Revista Asociación Española de Especialistas en Medicina del Trabajo, XXIX(1)*, 42-56.
- Arenal Gota, T., Viana Gárriz, J. L., & Belzunegui Otano, T. (2019). ¿Presenta desgaste profesional el personal de urgencias extrahospitalarias? Resultados Encuestas de Malasch. *Med Segur Trab, LXV*, 24-36.
- Attorresi, H. F., Lozzia, G. S., Abal, F. J., Galibert, M. S., & Aguerri, M. E. (Agosto de 2009). Teoría de respuesta al ítem: Conceptos básicos y aplicaciones para la medición de constructos psicológicos. *Revista Argentina de Clínica Psicológica, VIII(2)*, 179-188.
- Bernal García, M. I., Salamanca Jiménez, D. R., Perez Gutiérrez, N., & Quemba Mesa, M. P. (2018). Validez de contenido por juicio de expertos de un instrumento para medir percepciones físico-emocionales en la práctica de disección anatómica. *Educación Médica, XXI(6)*, 349-356.

Bolado, R., Ibáñez, J., & Lantarón, A. (1998). *El juicio de expertos*. Madrid, España: Consejo de Seguridad Nuclear.

Carrión García, M. Á., & Hernández Garcia, T. J. (2018). Factores psicosociales y desgaste psíquico en ámbito laboral. *Salud Uninorte, XXXIV(3)*, 705-714.

Castro Méndez, N. P. (2018). Riesgos Psicosociales y Salud Laboral en Centros de Salud. *Ciencia & Trabajo, XX(63)*, 155-159.

Castro, S. N., Infante, J. L., & Brust, M. E. (2014). Evaluación de factores psicosociales en el trabajo de la construcción: Avances y dificultades en la aplicación metodológica. *VIII Jornadas de Sociología de la UNLP* (págs. 1-21). Ensenada: Universidad Nacional de La Plata.

Charria O., V. H., Sarsosa P., K. V., & Arenas O., F. (2011). Factores de riesgo psicosocial laboral: Métodos e instrumentos de evaluación. *Revista Facultad Nacional de Salud Pública, XXIX(4)*, 380-391.

Cladellas Pros, R., Castelló Tarrida, A., & Parrado Romero, E. (2018). Satisfacción, salud y estrés laboral del profesorado universitario según su situación contractual. *Revista de Salud Pública, XX(1)*, 53-59.

Dahab, J., Rivadeneira, C., & Minici, A. (Marzo de 2010). El enfoque cognitivo-transaccional del estrés. *Revista de Terapia Cognitivo Conductual(18)*, 1-6.

de Vera Estévez, M. d. (1 de Mayo de 2004). *Módulo II: Moldes Cognitivos y Bienestar Subjetivo*. Obtenido de Los Moldes de La Mente:
<http://www.moldesmentales.com/otros/mar.htm>

Díaz Bambula, F., & Rentería Pérez, E. (2017). De la seguridad al riesgo psicosocial en el trabajo en la legislación colombiana de salud ocupacional. *Estudios Socio-Jurídicos, XIX(2)*, 129-155.

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, XII(6)*, 327-346.

- 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.
- Folkman, S., Lazarus, R. S., Gruen, R. J., & De Longis, A. (1986). Appraisal, Coping, Health Status, and Psychological Symptoms. *Journal of Personality and Social Psychology*, L(3), 571-579.
- Galeano Marín, M. E. (2009). *Estrategias de investigación social cualitativa: El giro de la mirada* (Segunda ed.). Medellín: La Carreta Editores.
- Galicia Alarcón, L. A., Balderrama Trápaga, J. A., & Edel Navarro, R. (Octubre de 2017). Validez de contenido por juicio de expertos: Propuesta de una herramienta virtual. *Apertura*, IX(2), 42-53.
- García Rubiano, M., & Forero Aponte, C. (2018). Estrés laboral y contrato psicológico como elementos relacionales del cambio organizacional. (U. C. Colombia, Ed.) *Diversitas - Perspectivas en Psicología*, XIV(1), 149-162.
- García Sánchez, J., Aguilera Terrats, J., & Castillo Rosas, A. (2011). Guía técnica para la construcción de escalas de actitud. *Odiseo: Revista Electrónica de Pedagogía*, VIII(16), 1-13.
- Gómez Benito, J., & Hidalgo Montesinos, M. D. (2003). Desarrollos recientes en psicometría. *Avances en Medición*(1), 17-36.
- Gómez, V., Segura Camacho, S., Castrillón, D., & Perilla, L. E. (2016). Estandarización de una batería para la evaluación de factores de riesgo psicosociales laborales en trabajadores colombianos. *Acta Colombiana de Psicología*, XIX(2), 221-238.
- González Romá, V. (1 de Enero de 2001). *Método de escalamiento unidimensional de Likert*. Obtenido de Universidad de Valencia: <https://www.uv.es/gonzalev/PSICOM-04-05/PSICOM-TEMA%2005-LIKERT.rtf>
- Gutiérrez Strauss, A. M., & Viloría Doria, J. C. (Abril de 2014). Riesgos psicosociales y estrés en el ambiente laboral. *Salud Uninorte*, XXX(1), 5-7.

- Hernández González, Á. I., Ortega Andeane, R. P., & Reidl Martínez, L. M. (2012). Validación del instrumento de estrés laboral para médicos mexicanos. *En-Claves del Pensamiento*, VI(11), 113-129.
- Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, P. (2014). *Metodología de la investigación* (Sexta ed.). México D.F., México: McGraw Hill/Interamericana Editores S.A.
- IBM (Dirección). (2013). *SPSS Statistics Base* [Película].
- JotForm Inc. (4 de Enero de 2021). *JotForm*. Obtenido de Constructor fácil de formularios en línea: <https://www.jotform.com/>
- Juárez Hernández, L. G., & Tobón, S. (2018). Análisis de los elementos implícitos en la validación de contenido de un instrumento de investigación. *Espacios*, XXXIX(53), 23-29.
- Krause, M. (Enero de 1995). La investigación cualitativa: Un campo de posibilidades y desafíos. *Temas de Educación*(7), 1-19.
- Lazarus, R. S., & Folkman, S. (1984). *Stress, Appraisal, and Coping* (Primera ed.). New York, New York, United States of America: Springer Publishing Company Inc.
- Lemos, M., Calle, G., Roldán, T., Valencia, M., Orejuela, J. J., & Román Calderón, J. P. (2019). Factores psicosociales asociados al estrés en profesores universitarios colombianos. *Diversitas: Perspectivas en Psicología*, XV(1), 61-72.
- Loaiza B., E., & Peña, A. E. (2013). Niveles de estrés y síndrome de Burnout en contadores públicos colombianos. *Actualidad Contable Faces*, XXVI(26), 27-44.
- Matas, A. (2018). Diseño del formato de escalas tipo Likert: Un estado de la cuestión. *Redie*, XX(1), 38-47.
- Meneses, J., Barrios, M., Bonillo, A., Cosculluela, A., Lozano, L. M., Turbany, J., & Valero, S. (2013). *Psicometría* (Primera ed.). Barcelona, Cataluña, España: UOC.
- Microsoft (Dirección). (2016). *Excel Office* [Película].

- Ministerio de La Protección Social. (17 de Julio de 2008). Resolución 2646. *Resolución 2646*. Bogotá D.C., Cundinamarca, Colombia: Gobierno de Colombia.
- Ministerio de La Protección Social. (2010). *Batería de Instrumentos para La Evaluación de Factores de Riesgo Psicosocial* (Tercera ed.). Bogotá D.C., Colombia: Pontificia Universidad Javeriana.
- Ministro de Salud y Protección Social. (10 de Noviembre de 2016). Decreto 1830. *Decreto Nacional*. Bogotá D.C., Cundinamarca, Colombia: Presidencia de la República.
- Morales Vallejo, P. (2010). *Guía para construir escalas de actitudes*. Madrid: Universidad Pontificia Comillas.
- Morales Vallejo, P. (2011). *Guía para construir cuestionarios y escalas de actitudes*. Guatemala: Universidad Rafael Landívar.
- Muñiz, J. (2010). Las Teorías de Los Tests: Teoría Clásica y Teoría de Respuesta a Los Ítems. *Papeles del Psicólogo*, XXXI(1), 57-66.
- Muñiz, J., & Fonseca Pedrero, E. (2019). Diez pasos para la construcción de un test. *Psicothema*, XXXI(1), 7-16.
- Muñoz Rojas, D., Orellano, N., & Hernández Palma, H. (1 de Julio de 2018). Riesgo psicosocial: Tendencias y nuevas orientaciones laborales. *Psicogente*, XXI(40), 532-544.
- Murillo Torrecilla, F. J. (2006). Cuestionarios y Escalas de Actitudes. *Formación de Profesorado y Educación*, 1-16.
- Navarro Saldaña, G., Flores Oyarzo, G., & González Navarro, M. G. (2020). Construcción y estudio psicométrico de un instrumento para evaluar inteligencia emocional en estudiantes chilenos. *Revista de estudios y experiencias en educación*, XIX(39), 29-43.
- Navas, M. J. (1994). Teoría Clásica de Los Tests versus Teoría de Respuesta al Ítem. *Psicológica*, XV, 175-208.
- Ortiz Barradas, A. M. (2019). Diseño y validación de una escala para intereses propedéuticos en México. *Ride*, X(19), 1-13.

- Pereira Pérez, Z. (2011). Los diseños de método mixto en la investigación en educación: Una experiencia concreta. *Educare*, XV(1), 15-29.
- Prada Ospina, R. (2019). Social psychological factors and their relation to work-related stress as generating effect of burnout. *Interdisciplinaria*, XXXVI(2), 39-53.
- Rega, L. L. (2010). Guía concisa para la elaboración de artículos científicos. *USAL. RedBUS*, 1-12.
- Ríos Uribe, D. (2014). *Tratamiento normativo de Los Factores de Riesgo Psicosocial Laborales en Colombia y su relación con El Acoso Laboral de La Resolución 2646 de 2008 y La Ley 1010 de 2006*. Medellín: Universidad de Antioquia.
- Robles Garrote, P., & Rojas, M. D. (2015). La validación por juicio de expertos: Dos investigaciones cualitativas en lingüística aplicada. *Revista Nebrija de lingüística aplicada a la enseñanza de las lenguas*, 1-16.
- Salamanca Velandia, S. R., Pérez Torres, J. M., Infante Alvarado, A. F., & Olarte Ardila, Y. Y. (2019). Análisis de los factores de riesgo psicosocial a nivel nacional e internacional. *Temas*, III(13), 39-45.
- Sarsosa Prowesk, K., & Charria Ortíz, V. H. (2018). Estrés laboral en personal asistencial de cuatro instituciones de salud nivel III de Cali, Colombia. *Universidad y Salud*, XX(1), 44-52.
- Souza Ramos, F. R., De Farias Brehmer, L. C., De Lima Dalmolin, G., Ramos Silveira, L., Ghizoni Schneider, D., & De Oliveira Vargas, M. A. (2020). Asociación entre distrés moral y elementos de apoyo de la deliberación moral en enfermeros. *Revista Latino-Americana de Enfermagem*(28), 1-8.
- Tristán López, A., & Molgado Ramos, D. (2007). *Tablas de Validez de Contenido*. San Luis Potosí, México: Instituto de Evaluación e Ingeniería Avanzada.
- Urrutia Egaña, M., Barrios Araya, S., Gutiérrez Núñez, M., & Mayorga Camus, M. (2014). Métodos óptimos para determinar validez de contenido. *Educación Médica Superior*, XXVIII(3), 547-558.

- Vásquez Gonzáles, G. C., Jiménez Macías, I. U., & Juárez Hernández, L. G. (2020). Construcción-validación del cuestionario sobre madurez de la gestión del conocimiento para la innovación educativa en universidades. *Apertura*, *XII*(1), 132-151.
- Verlanga Silvente, V., & Rubio Hurtado, M. (2012). Clasificación de pruebas no paramétricas. Cómo aplicarlas en SPSS. *REIRE*, *V*(2), 101-113.
- Vidal Lacosta, V. (2019). Estudio del estrés laboral en las PYMES en la provincia de Zaragoza. *Asociación Española de Especialistas en Medicina del Trabajo*, *XXVIII*(4), 254-266.
- Vieco Gómez, G. F., & Abello Llanos, R. (2014). Factores psicosociales de origen laboral, estrés y morbilidad en el mundo. *Psicología desde el Caribe*, *XXXI*(2), 354-385.
- Wright, V. (Productor), Stubbings, G., & Wright, V. (Dirección). (2006). *Mente biológica, muy estresada* [Película]. Inglaterra: TVE.
- Zamora de Ortiz, M. S., Serrano Pastor, F. J., & Martínez Segura, M. J. (2020). Validez de contenido del modelo didáctico P-VIRC mediante el juicio de expertos. *Formación Universitaria*, *XIII*(3), 43-54.

Anexos

Anexo A – Juicio de Expertos

JUICIO DE EXPERTOS

Revisión de ítems del Cuestionario para la Evaluación del Estrés (tercera versión) de la Bateria de Instrumentos para la Evaluación de Factores de Riesgo Psicosocial del Ministerio de Protección Social.

Juan Esteban Gallego Duque, con c.c., estudiante de Psicología de la Facultad de Ciencias Sociales y Humanas de la Universidad de Antioquia, con la asesoría del Profesor **Carlos Alberto Marín Monsalve**, con c.c. y Tarjeta Profesional #104480, se encuentran realizando el trabajo de grado: “Revisión de ítems del Cuestionario para la Evaluación del Estrés (tercera versión) de la Bateria de Instrumentos para la Evaluación de Factores de Riesgo Psicosocial del Ministerio de Protección Social”. Cuyo objetivo es evaluar los aspectos susceptibles de mejora en la formulación de los ítems, a partir de la identificación del modelo teórico usado para su construcción, y de la bibliografía existente alrededor del tema, centrándose únicamente en los ítems y sin pretender el desarrollo de un nuevo cuestionario.

Para dar cumplimiento con lo propuesto se realiza este Juicio de Expertos, como opinión informada de profesionales con trayectoria en el tema, reconocidos por otros profesionales como personas calificadas en la materia (por su formación y/o experiencia), que pueden brindar información, evidencia, juicios y valoraciones precisas.

Respetado Psicólogo/a, usted por su experiencia en temas de riesgo psicosocial y/o psicometría, hace parte del grupo de profesionales elegido para evaluar los ítems del Cuestionario para la evaluación del estrés (tercera versión), cuyos resultados harán parte de la Tesis de pregrado que será presentada a modo de Artículo, por consiguiente, la información por usted suministrada tendrá un uso netamente investigativo y podrá ser regulada en cualquier momento que lo considere oportuno.

Esta evaluación debe realizarse con máxima objetividad, para que los resultados sean válidos y eficientes en asuntos Psicométricos y de otras índoles relacionadas, por lo cual se aconseja llenar el cuestionario con tiempo, calma y plena disposición, y declinar del proceso en caso de considerar que dicha objetividad se encuentre comprometida por procesos o situaciones del pasado, relacionadas con el instrumento a evaluar o las Instituciones que participaron en su desarrollo.

Se agradece su valiosa participación, y cualquier aporte que pueda ayudar con el mejoramiento de este proceso será bien recibido.

Formación académica: _____

Cargo actual: _____

Institución/Empresa: _____

Nombres y Apellidos

Área (s) de Experiencia

A continuación, se presentarán dos grupos de ítems, el A que pertenece a los ítems originales del cuestionario, y el B que corresponde a los ítems dispuestos para reemplazar el contenido original con el fin de mejorar de manera general el instrumento. Es importante aclarar que los ítems del grupo A considerados correctos desde la Teoría del Ítem y las escalas tipo Likert, se transcribirán literalmente al grupo B, así mismo, los ítems del grupo A que tengan la posibilidad de ser divididos en más de una pregunta, aparecerán separados dentro del grupo B.

Solo se juzgará el cuadro de malestares, entorno a sus posibilidades de respuesta: Siempre; Casi Siempre; A Veces; Nunca.

Califique cada ítem de acuerdo con los siguientes indicadores:

CATEGORIA	CALIFICACION	INDICADOR
Suficiencia: Los ítems que pertenecen a una misma dimensión bastan para obtener la medición de ésta.	1. No cumple el criterio	Los ítems no son suficientes para medir la dimensión.
	2. Bajo nivel	Los ítems miden algún aspecto de la dimensión, pero no corresponden con la dimensión total.
	3. Moderado nivel	Se deben incrementar algunos ítems para poder evaluar la dimensión completamente.
	4. Alto nivel	Los ítems son suficientes.
Claridad: El ítem se comprende fácilmente, es decir, su sintáctica y semántica son adecuadas.	1. No cumple el criterio	El ítem no es claro.
	2. Bajo nivel	El ítem requiere bastantes modificaciones o una modificación muy grande en el uso de las palabras de acuerdo con su significado o por la ordenación de las mismas.
	3. Moderado	Se requiere una modificación muy

	nivel	específica de algunos de los términos del ítem.
	4. Alto nivel	El ítem es claro, tiene semántica y sintaxis adecuada.
Coherencia: El ítem tiene relación lógica con la dimensión o indicador que está midiendo.	1. No cumple el criterio	El ítem no tiene relación lógica con la dimensión.
	2. Bajo nivel	El ítem tiene una relación tangencial con la dimensión.
	3. Moderado nivel	El ítem tiene una relación moderada con la dimensión que está midiendo.
	4. Alto nivel	El ítem se encuentra completamente relacionado con la dimensión que está midiendo.
Relevancia: El ítem es esencial o importante, es decir, debe ser incluido.	1. No cumple el criterio	El ítem puede ser eliminado sin que se vea afectada la medición de la dimensión.
	2. Bajo nivel	El ítem tiene alguna relevancia, pero otro ítem puede estar incluyendo lo que mide éste.
	3. Moderado nivel	El ítem es relativamente importante.
	4. Alto nivel	El ítem es muy relevante y debe ser incluido.

Grupo A:

“Señale con una X la casilla que indique la frecuencia con que se le han presentado los siguientes malestares en los últimos tres meses”.

#	ITEMS (Fisiológicos)	SUFICIENCIA	COHERENCIA	RELEVANCIA	CLARIDAD	OBSERVACIONES
1	Dolores en el cuello y espalda o tensión muscular.					
2	Problemas gastrointestinales, úlcera péptica, acidez, problemas digestivos o del colon.					
3	Problemas respiratorios.					
4	Dolor de cabeza.					
5	Trastornos del sueño como somnolencia durante el día o desvelo en la noche.					
6	Palpitaciones en el pecho o problemas cardíacos.					
7	Cambios fuertes del apetito.					
8	Problemas relacionados con la función de los órganos genitales (impotencia, frigidez).					
#	ITEMS (Comportamiento Social)	SUFICIENCIA	COHERENCIA	RELEVANCIA	CLARIDAD	OBSERVACIONES
9	Dificultad en las relaciones familiares.					

10	Dificultad para permanecer quieto o dificultad para iniciar actividades.					
11	Dificultad en las relaciones con otras personas.					
12	Sensación de aislamiento y desinterés.					
#	ITEMS (<i>Intelectuales Y Laborales</i>)	SUFICIENC IA	COHERENC IA	RELEVANC IA	CLARIDA D	OBSERVACIO NES
13	Sentimiento de sobrecarga de trabajo.					
14	Dificultad para concentrarse, olvidos frecuentes.					
15	Aumento en el número de accidentes de trabajo.					
16	Sentimiento de frustración, de no haber hecho lo que se quería en la vida.					
17	Cansancio, tedio o desgano.					
18	Disminución del rendimiento en el trabajo o poca creatividad.					
19	Deseo de no asistir al trabajo.					
20	Bajo compromiso o poco interés con lo que se hace.					
21	Dificultad para tomar decisiones.					
22	Deseo de cambiar de empleo.					

#	ITEMS (Psicoemocional es)	SUFICIENC IA	COHERENC IA	RELEVANC IA	CLARIDA D	OBSERVACIO NES
23	Sentimiento de soledad y miedo.					
24	Sentimiento de irritabilidad, actitudes y pensamientos negativos.					
25	Sentimiento de angustia, preocupación o tristeza.					
26	Consumo de drogas para aliviar la tensión o los nervios.					
27	Sentimientos de que "no vale nada", o "no sirve para nada".					
28	Consumo de bebidas alcohólicas o caféo cigarrillo.					
29	Sentimiento de que está perdiendo la razón.					
30	Comportamiento srigidos, obstinación o terquedad.					
31	Sensación de no poder manejar los problemas de la vida.					

Grupo B:

“Señale con una X la casilla que indique la frecuencia con que se le han presentado los siguientes malestares en los últimos tres meses”.

#	ITEMS (Fisiológicos)	SUFICIENCIA	COHERENCIA	RELEVANCIA	CLARIDAD	OBSERVACIONES
1	Dolores lumbares.					
1.1	Tensión muscular.					
2	Problemas gastrointestinales.					
2.1	Problemas digestivos.					
3	Problemas respiratorios.					
4	Dolor de cabeza.					
5	Trastornos del sueño.					
5.1	Somnolencia durante el día.					
6	Problemas cardíacos.					
6.1	Taquicardias.					
7	Cambios fuertes en el apetito.					
8	Disminución del apetito sexual.					
#	ITEMS (Comportamiento Social)	SUFICIENCIA	COHERENCIA	RELEVANCIA	CLARIDAD	OBSERVACIONES
9	Dificultad en las relaciones familiares.					
10	Dificultad para permanecer quieto.					
10.1	Dificultad para iniciar actividades.					
11	Dificultad en las relaciones interpersonales.					
12	Sensación de aislamiento.					
12.1	Desinterés.					

#	ITEMS (<i>Intelectuales Y Laborales</i>)	SUFICIENC IA	COHERENC IA	RELEVANC IA	CLARIDA D	OBSERVACIO NES
13	Sentimiento de sobrecarga de trabajo.					
14	Dificultad para concentrarse.					
14.1	Olvidos frecuentes.					
15	Aumento en el número de accidentes laborales.					
16	Sentimiento de no haber hecho lo que se quería.					
16.1	Frustración.					
17	Sensación de desgaste.					
17.1	Sensación de tedio.					
18	Disminución del rendimiento en el trabajo.					
18.1	Falta de creatividad.					
19	Desánimo para asistir al trabajo.					
20	Poco interés con lo que se hace.					
20.1	Bajo compromiso con lo que se hace.					
21	Dificultad para tomar decisiones.					
22	Cuestionamientos sobre cambio de empleo.					
#	ITEMS (<i>Psicoemocional es</i>)	SUFICIENC IA	COHERENC IA	RELEVANC IA	CLARIDA D	OBSERVACIO NES
23	Sentimiento de soledad.					
23.1	Sensación de miedo.					
24	Sentimiento de irritabilidad.					
24.1	Actitudes negativas.					

24.2	Pensamientos negativos.					
25	Sentimiento de angustia.					
25.1	Sensación de preocupación.					
25.2	Tristeza.					
26	Consumo de drogas.					
27	Sentimiento de "no servir para nada".					
28	Consumo frecuente de bebidas alcohólicas.					
29	Sensación de estar perdiendo la razón.					
30	Comportamiento obstinados frecuentes.					
31	Sensación de no poder manejar los problemas de la vida.					

Fecha: / /

FIRMA

Bibliografía

- Abad, F. J., Garrido, J., Olea, J., & Ponsoda, V. (2006). *Introducción a la Psicometría: Teoría clásica de los tests y teoría de la respuesta al ítem*. Madrid, España: Universidad Autónoma de Madrid.
- Escobar Pérez, J., & Cuervo Martínez, Á. (2008). Validez de contenido y juicio de expertos: Una aproximación a su utilización. *Avances en Medición*, 27-36.
- Galicia Alarcón, L. A., Balderrama Trápaga, J. A., & Edel Navarro, R. (Octubre de 2017). Validez de Contenido por Juicio de Expertos: Propuesta de una herramienta virtual. *Apertura*, IX(2), 42-53.
- González Romá, V. (1 de Enero de 2001). *Método de escalamiento unidimensional de Likert*. Obtenido de Universidad de Valencia: <https://www.uv.es/gonzalev/PSICOM-04-05/PSICOM-TEMA%2005-LIKERT.rtf>
- Ministerio de La Protección Social. (2010). *Batería de Instrumentos para La Evaluación de Factores de Riesgo Psicosocial* (Tercera ed.). Bogotá D.C., Colombia: Pontificia Universidad Javeriana.
- Muñiz, J. (2010). Las Teorías de Los Tests: Teoría Clásica y Teoría de Respuesta a Los Ítems. *Papeles del Psicólogo*, XXXI(1), 57-66.
- Robles Garrote, P., & Rojas, M. D. (2015). La validación por juicio de expertos: Dos investigaciones cualitativas en Lingüística aplicada. *Nebrija*, 1-16.

Anexo B - Respuestas del Juicio de Expertos

JUEZ	SEXO	CAMPO OCUPACIONAL	NIVEL PROFESIONAL
1	M	P	Doctorado
2	F	SST	Especialización
3	F	SST	Especialización
4	F	SST	Especialización
5	F	P	Maestría
6	F	RP	Especialización
7	M	P	Maestría
8	F	SST	Especialización
9	F	RP	Maestría
10	F	RP	Especialización