

Motivos que han propiciado que la evaluación de riesgos psicosociales sea la asignatura pendiente de la prevención de riesgos laborales. Estudio Delphi

Reasons that have made psychosocial risk assessment the pending subject of occupational risks prevention. Delphi study

Rita Louzán Mariño  | rita.louzan@unir.net
Universidad Internacional de La Rioja, España

10.17502/mrcs.v10i2.556

Recibido: 03-05-2022
Aceptado: 27-07-2022



Resumen

A pesar de ser etiquetados como la epidemia del siglo XXI, los riesgos psicosociales apenas son evaluados por las organizaciones. La ausencia de una cultura preventiva, el escaso control de fuentes de error en las mediciones o la falta de formación específica podrían explicar la dificultad de evaluar estos riesgos. En este sentido, el principal objetivo de esta investigación estriba en sacar a la luz las posibles debilidades que entorpecen las evaluaciones de riesgos psicosociales de origen laboral, y con ello poder mejorar la prevención de la salud mental y física de los trabajadores expuestos al estrés laboral. Se ha realizado consulta directamente a un grupo de reputados académicos y profesionales expertos en la materia, utilizando la metodología Delphi. Los resultados muestran, por una parte, un elevado consenso sobre las principales fuentes de amenaza. Por otra parte, dichos expertos sugieren una serie de directrices prácticas. Como conclusión son identificadas, según el juicio experto, variedad de amenazas que debilitan las evaluaciones psicosociales, a la vez que se recomiendan estrategias de mejora para realizar las evaluaciones de riesgos psicosociales.

Palabras clave: Delphi, estrés laboral, fuentes de error, riesgos psicosociales, validez.

Abstract

Despite being labeled as the epidemic of the 21st century, psychosocial risks are hardly evaluated by organizations. The absence of a preventive culture, the limited control of sources of error in measurements or the lack of specific training could explain the difficulty of assessing these risks. In this sense, the main objective of this research is to bring to light the possible weaknesses that hinder the evaluations of psychosocial risks of work origin, and with this to be able to improve the prevention of mental and physical health of workers exposed to work stress. A group of renowned academics and experts in the field was consulted directly, using the Delphi methodology. The results show, on the one hand, a high consensus on the main sources of threat. On the other hand, these experts suggest a series of practical guidelines aimed at strengthening these evaluations. In conclusion, according to expert judgment, a variety of threats that weaken psychosocial evaluations are identified, while improvement strategies are recommended to carry out psychosocial risk evaluations.

Keywords: Delphi, error sources, occupational stress, psychosocial risk at work, validity.

Sumario

1. Introducción | 2. Metodología | 2.1. Formulación del problema | 2.2. Selección del panel de expertos | 2.3. Confección y envío de cuestionarios | 2.4. Análisis de datos | 3. Resultados | 3.1. Ítems cuantitativos | 3.2. Ítems cualitativos | 4. Discusión y conclusiones | Referencias.

Cómo citar este artículo

Louzán, R. (2022). Motivos que han propiciado que la evaluación de riesgos psicosociales sea la asignatura pendiente de la prevención de riesgos laborales. Estudio Delphi. *methaodos.revista de ciencias sociales*, 10(2): 226-242. <http://dx.doi.org/10.17502/mrcs.v10i2.556>

1. Introducción

Los factores psicosociales son condiciones psicológicas del trabajo que como tales pueden ser positivas y negativas. De acuerdo con Moreno y Báez (2010), cuando las condiciones son adecuadas aumentan los niveles de motivación y satisfacción laboral, afectando positivamente a la salud. En cambio, cuando adoptan una configuración adversa —por ejemplo, exposición a ritmos de trabajo excesivo, trato injusto, inseguridad o falta de apoyo social—, se consideran factores psicosociales de riesgo (estresores) porque provocan respuestas de inadaptación y de tensión en los trabajadores, manifestándose con mecanismos psicológicos y fisiológicos donde opera el estrés como precursor de la enfermedad (Gil-Monte, 2010).

En este sentido, existe abundante literatura científica que señala al estrés laboral como principal responsable de infinidad de daños a la salud de los trabajadores (Moreno y Báez, 2010). Entre estos trastornos destacan por su incidencia la ansiedad y la depresión, llegando incluso a provocar el suicidio (INSST, 2018). No en vano, a principios del año 2000 han sido identificados los riesgos psicosociales como uno de los grandes males del siglo XXI (Baratech, 2006).

Los efectos del estrés no se circunscriben únicamente a la salud de las personas, también tienen efectos nocivos en las propias organizaciones. Según la Inspección de Trabajo y Seguridad Social (ITSS, 2021), se presume que el estrés laboral es responsable del 50% del absentismo y de 1/5 de la rotación de personal, además de provocar reducción del rendimiento e incremento de la siniestralidad.

A pesar de su evidente relevancia, apenas se tienen en cuenta los riesgos psicosociales en su planificación preventiva, así se desprende de la reciente ESENER-3 (EU-OSHA, 2019), donde los riesgos psicosociales constituyen uno de los mayores problemas reportados. Llama especialmente la atención que un 61,4% de empresas españolas no poseen ningún plan de gestión para prevenir el estrés laboral, lo cual actúa en detrimento de la salud física y mental de los trabajadores.

Una de las debilidades de las evaluaciones de riesgos psicosociales puede estar relacionada con la controversia generada por la falta de legislación específica que arroje luz sobre cómo evaluar dichos riesgos (Velázquez, 2020). Esto es, porque a diferencia de otros riesgos, en España no existen disposiciones legales específicas, ni normas UNE, ni directrices precisas de organismos oficiales. Ante esta realidad, se deben utilizar guías de entidades de reconocido prestigio en la materia u otros métodos o criterios profesionales descritos documentalmente, siempre y cuando proporcionen confianza en el resultado (ITSS, 2012; R.D. 39/1997), lo que ha contribuido a la proliferación de métodos de evaluación de muy distinta calidad, fundamentalmente basados en autoinformes, pues son de menor coste, más rápidos y flexibles en su aplicación, uso y análisis.

El inconveniente es que no existe un consenso profesional sobre las herramientas a utilizar, y pocas han demostrado contar con propiedades psicométricas satisfactorias (Moreno y Báez, 2010). De ellas, tan solo el Método CoPsoQ ISTAS 21 (Moncada *et al.*, 2021) y el F-PSICO (Ferrer, *et al.*, 2011) han sido baremadas con una muestra amplia de población española.

Con todo, no basta contar con herramientas fiables y válidas, es preciso hacer un uso profesional de ellas y tratar de minimizar las posibles fuentes de error (Louzán, 2020). Por ejemplo, si no se garantiza el anonimato y la confidencialidad de los datos o si se altera el instrumento de medida, es muy posible que se obtengan datos sesgados que pondrán en entredicho la validez de los resultados.

Por tanto, es esencial contar con profesionales expertos y con formación específica en evaluar riesgos psicosociales, pues evaluar este tipo de riesgos va mucho allá de administrar e interpretar los resultados de un cuestionario (INSST, 2022; Oncins y Almodovar, 1997). Sin embargo, en España existe una muy baja demanda de técnicos de prevención especializados en el área psicosocial (15,70%), en comparación con profesionales de las otras especialidades, los cuales se demandan en más del 75% (EU-OSHA, 2015).

En definitiva, la falta de legislación específica, la falta de consenso sobre el uso de metodologías adecuadas, su incorrecta aplicación o el déficit de profesionales cualificados han configurado lo que Niño (2006, p.30) denomina la “ceremonia de la confusión” en el ámbito de la Ergonomía y Psicología Aplicada. Dicha confusión podría ser la causante de que la evaluación de los riesgos psicosociales en el trabajo sea considerada como la asignatura pendiente de la prevención.

A tal efecto, la presente investigación surge debido a la falta de trabajos científicos que examinen específicamente y detalladamente las dificultades inherentes a las evaluaciones de riesgo en el ámbito de la Psicología Aplicada a nivel nacional, en particular desde una aproximación cualitativa. Concretamente, este trabajo tiene como principal objetivo conocer el juicio experto acerca de las principales debilidades que afectan a las evaluaciones de riesgo psicosocial en España.

De ese modo, se diseñó un estudio Delphi en el que 20 expertos del ámbito académico y profesional fueron invitados a tomar parte en una encuesta. El grado de consenso obtenido y las posibles opiniones divergentes permitirán componer un escenario real sobre las principales debilidades o amenazas que atañen a las evaluaciones de riesgos psicosociales y, a su vez, permitirán identificar acciones dirigidas a mejorar su validez.

2. Metodología

La presente investigación se ha basado en la aplicación de método Delphi (Dalkey y Helmer, 1963) modificado a dos rondas, con objeto de obtener una visión experta sobre los principales retos a los que se enfrentan los académicos y profesionales implicados en la evaluación de riesgos psicosociales. El Delphi empleado abarca 4 fases principales.

2.1. Formulación del problema

Se trató de definir el problema de investigación en base a la revisión teórica del estado de la cuestión. Para ello, fueron consultadas numerosas fuentes especializadas, como libros, revistas, legislación, recomendaciones del INSST y de la ITSS, además de explorar diversas bases de datos bibliográficas: *Medline*, *ProQuest*, *PsycInfo*, *Science Direct* *Scifinder* y *SCOPUS*. Fruto de esta exploración, se han extraído las principales cuestiones referidas a la problemática de evaluar riesgos psicosociales. Se hizo especial hincapié en el uso de medidas de autoinforme, pues según la literatura, forman parte de la polémica. De este modo se han articulado los principales ejes de discusión y, a partir de ellas, se ha diseñado la formulación del primer cuestionario con ítems cerrados y abiertos.

2.2. Selección del panel de expertos

Para formar parte de dicho panel, se han escogido por un lado académicos del ámbito de la Psicología con actividad investigadora en riesgos psicosociales, y, por otro, técnicos de prevención en activo, especializados en la prevención de riesgos psicosociales, ambos con conocimiento y experiencia en la temática a nivel nacional.

La parte más importante de todo el proceso es la relativa a la selección de los expertos, pues en este tipo de estudios la credibilidad y la fiabilidad la otorga la calidad del grupo seleccionado (Okoli y Pawlowski, 2004). Intentando efectuar una selección con la mayor rigurosidad posible, se ha aplicado un doble enfoque: cualitativo y cuantitativo.

A. Enfoque cualitativo

Atendiendo a la calidad de los participantes, se ha realizado la selección según la facultad del sujeto de aportar información relevante a la investigación. Se han utilizado requisitos de selección muy rigurosos, como es la combinación de criterios de trayectoria académica y profesional (biograma) con la autoevaluación del propio experto (*self ratings*) sobre sus conocimientos en la temática a tratar (Cabero y Llorente, 2013; López-Gómez, 2018). En este estudio se ha utilizado como criterio autoevaluativo el *Coficiente de Competencia de Experto (K)* (fig. 1), cuya formulación es la siguiente:

$$K = \frac{1}{2} (K_c + K_a)$$

Esta metodología está fundamentada en la cualimetría elaborada por el Comité Estatal para la Ciencia y la Técnica de Rusia en el año 1971. Se emplea para determinar el coeficiente de competencia de un candidato como experto, siendo "1" el valor máximo (Burguet *et al.*, 2019) (Figura 1).

Figura 1. Autoevaluación del Criterio experto

Kc Es el *Coefficiente de conocimiento* que se posee sobre el tema planteado. Para su cálculo el experto asignará un valor que, a su juicio, represente lo más fielmente posible el grado de información que tiene sobre el asunto a tratar, dentro de una escala de 0 a 1. Donde 0 representa el desconocimiento total y 10 el pleno conocimiento en la materia, el resultado se multiplica por 0.1

0	1	2	3	4	5	6	7	8	9

Ka Corresponde al *Coefficiente de argumentación*. Para hallar su valor el experto debe indicar el grado de influencia que considera que han tenido cada una de las seis fuentes indicadas en la tabla, en la fundamentación de sus criterios acerca de la materia objeto de debate. A cada grado marcado le corresponde la puntuación siguiente:

		Grado de influencia de cada una de las fuentes en sus criterios		
FUENTES		A (alto)	M (medio)	B (bajo)
1º	Análisis teóricos realizados por usted	.30	.20	.10
2º	Su experiencia en el tema	.50	.40	.20
3º	Trabajos de autores nacionales consultados	.05	.05	.05
4º	Trabajos de autores extranjeros consultados	.05	.05	.05
5º	Su propio conocimiento del estado del problema en el extranjero	.05	.05	.05
6º	Su formación	.05	.05	.05

Fuente. Elaboración propia.

El criterio fijado para la presente investigación ha sido un exigente $K \geq 0.8$ (Marín-González *et al.* 2021).

B. Enfoque cuantitativo

Según la literatura de referencia, el número de sujetos idóneos para conformar el panel se encontraría entre 7 y 35 participantes (López-Gómez, 2018). Esto es así porque los primeros estudios de Dalkey (1972) mostraron que el error disminuía de manera significativa por cada participante añadido hasta llegar a 7, y la curva de saturación a partir de 30 individuos no mejoraba de manera proporcional la calidad de los datos obtenidos.

Atendiendo a estas premisas, se ha querido conseguir una muestra encuadrada en dichos parámetros. Se comenzó realizando una búsqueda de psicólogos con publicaciones originales sobre el tema de estudio en las bases de datos bibliográficas PsycInfo, Science Direct y Dialnet. Para la identificación de técnicos de prevención especializados en riesgos psicosociales, se utilizó la técnica de *Snowball sampling* (Hernández-Sampieri y Mendoza, 2018). A continuación, se confeccionó para cada uno de los posibles participantes identificados un biograma basado en la información obtenida de sus perfiles públicos. De este modo, y haciendo uso de sus direcciones públicas de correo electrónico, se invitó a participar a 117 expertos.

De todos ellos, aceptaron participar 20 expertos (17.00% de los invitados). Un 15% declinaron participar por falta de tiempo y el resto no contestaron a la propuesta. Al utilizar como medio de contacto direcciones de email halladas por internet, no existe la certeza de que dichos correos electrónicos se encuentren actualizados y, por consiguiente, no se puede asegurar que todas las invitaciones hayan llegado a sus destinatarios. Una vez aplicado el tamiz del *self rating* ($K \geq 0.8$), la muestra quedó compuesta por un total de 16 miembros.

2.3. Confección y envío de los cuestionarios

Sobre el número de rondas necesario, tampoco existe un consenso establecido. No obstante, utilizar un número excesivo de rondas para intentar reducir la inestabilidad o la falta de convergencia de opiniones, dilatará desmesuradamente la duración del estudio, pudiendo fomentar el cansancio de los panelistas e incitarlos al abandono o, lo que es peor, a un consenso forzado (Brill *et al.*, 2006). Con todo, los mayores cambios en las respuestas Delphi ocurren en las dos primeras rondas, siendo este formato el más utilizado (Steurer, 2011). Siguiendo estos argumentos, y para garantizar la estabilidad y la continuidad del panel, se

establecieron de antemano dos rondas, distribuidas mediante correo electrónico entre noviembre del 2015 y de marzo del 2016.

En el cuerpo del mensaje de la primera ronda, se realizó un escrito de presentación informando del objetivo de la investigación y de las instrucciones para rellenar el test.

Los cuestionarios utilizados en ambas rondas fueron confeccionados garantizando el anonimato de cada participante, incluyéndose ítems tanto cualitativos como cuantitativos. Los ítems cuantitativos utilizados son de escalamiento tipo Likert de 5 niveles, con los siguientes calificadores lingüísticos: totalmente en desacuerdo (1), en desacuerdo (2), indeciso (3), de acuerdo (4) y totalmente de acuerdo (5).

Primera Ronda (R1)

La primera vuelta consistió en el envío del primer cuestionario a cada uno de los 117 expertos seleccionados. El cuestionario enviado en la R1 estaba compuesto por 32 ítems (Tabla 1)¹ cuidadosamente contruidos para evitar sesgos (25 ítems cerrados y 7 ítems abiertos). En el cuerpo del mensaje se realizó un escrito de presentación, informando del objetivo de la investigación y de las instrucciones para rellenar el test.

Los ítems cerrados fueron generados a partir de la revisión narrativa y los ítems abiertos se formularon de modo que los expertos pudiesen introducir otras variables no observadas en el cuestionario. De acuerdo con su contenido, el cuestionario fue dividido en 4 apartados:

1. Afirmaciones generales
2. Diseño de los instrumentos de medida que evalúan riesgos psicosociales
3. Modo de aplicación de estos instrumentos; y, por último
4. Corrección e interpretación de los resultados.

Además de los ítems objeto de estudio, se incluyó una pequeña batería de preguntas de autoevaluación destinada a conocer el Coeficiente de competencia de experto (K).

Una vez recibidas las respuestas al primer cuestionario, se confeccionó la lista definitiva de panelistas, atendiendo al filtro de grado de experto alcanzado (K). Cumpliendo esta premisa, el panel quedó definitivamente conformado por los 16 expertos.

Segunda Ronda (R2)

En vista de los resultados alcanzados en la R1, el propósito de la segunda vuelta fue promover el consenso, obteniendo una menor dispersión de respuestas. Con ese fin, se devolvió el cuestionario a los 16 panelistas participantes, informándoles de los resultados alcanzados en la R1 con objeto de ofrecer retroalimentación y promover así un mayor consenso. Los expertos, a la luz de los resultados de la primera ronda, pudieron reconsiderar y modificar sus respuestas. Este cuestionario siguió la misma estructura que el primero a excepción de las preguntas abiertas, cuyas respuestas se transformaron en 32 ítems cuantitativos y fueron incorporadas a modo de sumario, de modo que todos los participantes pudiesen conocer las aportaciones realizadas y valorarlas nuevamente según el formato Likert de 5 puntos descrito anteriormente.

La participación en esta ronda resultó mejor de lo esperado, al obtener una nula tasa de abandonos (0%), lo que refuerza la teoría de que un menor número de rondas favorece la motivación de los panelistas y reduce los índices de abandono (Brill *et al.*, 2006), avalando la decisión de utilizar un Delphi a dos vueltas.

2.4. Análisis de datos

El tratamiento de los datos siguió un proceso riguroso y meticuloso con el fin de conseguir interpretaciones exentas de la máxima subjetividad posible. Según esto, se diseñó el tratamiento estadístico de los ítems cuantitativos para obtener, por un lado, un análisis de los niveles de consenso en cada una de las iteraciones y, por otro lado, para establecer la estabilidad entre rondas.

¹ El cuestionario empleado tiene como objeto conocer la opinión individual de cada experto, como no se trata de ninguna escala de medida no se realiza el análisis de la consistencia interna.

Se han utilizado las medidas típicas de este tipo de estudios (López-Gómez, 2018; Von der Gracht, 2012): la mediana (Me) como medida de tendencia central, los cuartiles como normas de posición (Q1, Q2, Q3), el rango intercuartílico relativo (RIR) como índice de consenso, y el rango intercuartílico (RI) y el coeficiente de variación (CV) como medidas de dispersión.

Tabla 1. Batería de ítems de la encuesta Delphi

Afirmaciones generales (Apartado I)	
A1	La mayoría de empresas concede más importancia a la evaluación de los riesgos físicos que a la evaluación de los riesgos psicosociales
A2	Son pocas las empresas que cuentan con procedimientos para gestionar riesgos psicosociales tales como estrés, <i>mobbing</i> , <i>burnout</i> , violencia, etc.
A3	A menudo se confunde lo que es un método de evaluación de riesgos (método científico), con el hecho de administrar cuestionarios que evalúan riesgos (técnica de recogida de datos y una de las fases de la metodología)
A4	La evaluación de riesgos implica más procesos que valorar los riesgos mediante un cuestionario
Diseño de instrumentos de medida (Apartado II)	
A5	Para valorar los riesgos psicosociales es esencial la autopercepción del trabajador
A6	Para que los datos recogidos sean lo más objetivos posible, la herramienta a utilizar para la recogida de datos debería contener criterios que cumplan las propiedades de fiabilidad y validez
A7	Para que los datos recogidos sean lo más objetivos posible, la herramienta a utilizar para la recogida de datos debería contener criterios que controlen los sesgos
A8	Para que los datos recogidos sean lo más objetivos posible, la herramienta a utilizar para la recogida de datos debería contener criterios de baremación y estandarización
A9	Para que los datos recogidos sean lo más objetivos posible, la herramienta a utilizar para la recogida de datos debería contener fundamentación teórica
A10*	Otros criterios no indicados en el cuestionario que debería contener la herramienta de recogida de información para ser lo más objetivos posible
A11	Es poco apropiado incluir en los cuestionarios de evaluación de riesgos psicosociales escalas cuya finalidad no es detectar riesgos, como son escalas de satisfacción o clima laboral
Aplicación (Apartado III)	
A12	Las medidas de autoinforme son susceptibles de ser falseadas
A13	A la hora de aplicar un test suele darse más importancia a las propiedades psicométricas de fiabilidad y validez que a la detección de sesgos
A14	La aquiescencia es uno de los sesgos que pueden atribuirse a escalas de autoinforme utilizadas en prevención de riesgos psicosociales
A15	La deseabilidad social es uno de los sesgos que pueden atribuirse a escalas de autoinforme utilizadas en prevención de riesgos psicosociales
A16	El autoengaño o la simulación son sesgos que pueden atribuirse a escalas de autoinforme utilizadas en prevención de riesgos psicosociales
A17*	Otros sesgos no indicados en el cuestionario que puedan atribuirse a escalas de autoinforme utilizadas en prevención de riesgos psicosociales
A18	Un protocolo de administración donde, por ejemplo, los encargados de administrar los test lean las instrucciones pausada y claramente y traten de reducir la ansiedad de los evaluados, contribuiría a minimizar los posibles sesgos de respuesta
Corrección e Interpretación de resultados (Apartado IV)	
A19	A la hora de administrar test de evaluación de riesgos psicosociales no es frecuente realizar control de posibles sesgos o fuentes de error
A20	Los sesgos son un problema importante pues podrían contaminar los resultados de medida de los test, pudiendo incluso invalidar toda la evaluación
A21	Los sesgos de respuesta son posiblemente los de mayor incidencia en los cuestionarios de autoinforme
A22	Actualmente apenas existen herramientas para evaluar riesgos psicosociales que contengan control de sesgos
A23	A pesar de existir técnicas destinadas a detectar y minimizar los sesgos de respuesta a menudo son eludidos
A24	La percepción de incumplimiento de contrato psicológico, por parte del trabajador, podría influir en el falseamiento consciente de las respuestas de escalas utilizadas en la evaluación de riesgos psicosociales
A25	Es necesario tener conocimientos de psicometría para interpretar los resultados de las escalas de autoinforme utilizadas para valorar los riesgos psicosociales

*Ítems cualitativos

Fuente. Elaboración propia.

Para discriminar los reactivos con consenso de aquellos que no alcanzasen acuerdo o resultasen indeterminados, se utilizaron los dos criterios establecidos en Boix *et al.* (2008). En primer lugar, se empleó el porcentaje de acuerdo según la posición de la mediana (Me). Para ello, la escala de valoraciones se agrupó en tres intervalos (C1 y C2), (C3), y (C4 y C5), y, según la bibliografía revisada, se considera que existe consenso cuando los expertos que puntuasen fuera de la región de tres puntos que contiene la mediana fuesen menos de la tercera parte del total de los encuestados. En ese caso sería la posición de la mediana la que determina el acuerdo. Se consideraría consenso "favorable" cuando $Me \geq 4$; consenso en "desfavorable" si $Me \leq 2$; los casos en los que $Me > 2$ y < 4 se considerarían "indeterminados". En segundo lugar, se empleó el Rango Intercuartílico Relativo (RIR), donde un $RIR > .50$ implicaría falta de consenso. Como refuerzo de estos índices se aplicó también el Coeficiente de Variación (CV), considerando que existiría baja variación de respuesta si no superaban el valor de .26 (Ferrán, 2001). En la Tabla 2 se recopilan los criterios empleados para establecer la falta de consenso.

Tabla 2. Criterios establecidos del grado de no consenso

Orden	Criterios de Decisión	Resultado
1º	$Me > 2 < 4$	Indeterminado: (ni acuerdo, ni desacuerdo)
2º	$RIR \geq 0.50$	No consenso
3º	% de respuestas fuera del rango donde se sitúa la Me $> 33\%$	No consenso
Orden	Criterios de Decisión	Resultado
4º	$CV \geq 0.26$ (26%)	No consenso por exceso de variabilidad de respuesta

Me = Mediana / RIR = Rango Intercuartílico Relativo / CV = Coeficiente de Variación

Fuente. Elaboración propia.

Además del consenso, al final de las iteraciones se exploró el grado de estabilidad entre rondas. Independientemente del grado de convergencia alcanzado, se entendería por estabilidad la no variación estadísticamente significativa de los juicios de los expertos (Landeta, 2002). Para su exploración se utilizaron como indicadores de medida la variación del RIR (VRIR) y del CV (\cdot). Los baremos de decisión de estabilidad se establecieron utilizando criterios de mayor exigencia que los utilizados por otros estudios (p.e., Von der Gracht, 2012), asumiendo estabilidad cuando $-.25 \leq VRIR \leq .25$ y $\cdot \leq .27$. Se considera estabilidad satisfactoria si las observaciones superan el 80% .

Respecto a las respuestas de las preguntas abiertas recabadas en la primera ronda, éstas se presentaron convertidas en 32 ítems cerrados en el segundo cuestionario. Cada uno de los ítems se presentaron ordenados de acuerdo con los siguientes criterios: primero se ordenaron de forma decreciente de acuerdo con el valor de la media aritmética y, en segundo lugar, se ordenaron por el valor del RIR. Asimismo, se complementó la información con los porcentajes de acuerdo, tanto a favor como en contra, y con la mediana (Me). Se categorizarían estas cuestiones como relevantes si combinaban valores de preferencia y consenso con alta significación; es decir, si cumplían los siguientes criterios: $Me \neq 3$, frecuencia de respuesta $\geq 75\%$, y $RIR \leq 0.5$.

3. Resultados

3.1. Ítems cuantitativos

Los resultados de los ítems sobre los que se alcanzó consenso se muestran en la Tabla 3. En la primera ola se observaron 15 ítems en los que el grado de acuerdo fue máximo. De hecho, la mediana fue para todos estos ítems ≥ 4 ; es decir, el sentido del consenso "a favor".

En la 2ª ronda del estudio Delphi los mismos panelistas volvieron a reflexionar sobre todos los ítems, incluso aquellos que alcanzaran consenso en la R1. En esta nueva ola, se redujo notablemente el porcentaje

de respuestas dudosas o de indecisión (C3) y se incrementó el grado de acuerdo favorable en todos y cada uno de los ítems ($Me \geq 4$), a excepción de los ítems A13) y A20). Respecto al ítem A15, se ha hallado una tasa de acuerdo a favor del 68.75% y un RIR $\leq .50$ (superando el consenso mínimo exigido), si bien sólo sobrepasa tímidamente los límites de dispersión fijados.

Tabla 3. Resultados de consenso R1 y R2 (n=16)

RONDA 1							RONDA 2						
Ítem	Me	RIR	CV	% A favor	% En contra	Grado	Ítem	Me	RIR	CV	% A favor	% En contra	Grado
A1	5.00	0.15	9.42	100.00	0.00	A	A1	5.00	0.00	0.08	100.00	0.00	U
A2	4.00	0.44	25.73	68.75	12.50	S	A2	4.00	0.25	0.16	87.50	0.00	A
A3	5.00	0.20	16.23	87.50	0.00	A	A3	5.00	0.00	0.08	100.00	0.00	A
A4	5.00	0.00	8.38	100.00	0.00	A	A4	5.00	0.00	0.05	100.00	0.00	A
A5	4.50	0.39	20.15	75.00	0.00	S	A5	4.00	0.25	0.16	87.50	0.00	A
A6	5.00	0.20	14.05	93.75	0.00	A	A6	5.00	0.00	0.08	100.00	0.00	A
A7	5.00	0.20	14.05	93.75	0.00	A	A7	5.00	0.20	0.11	100.00	0.00	A
A8	4.50	0.22	16.43	87.50	0.00	A	A8	5.00	0.20	0.14	93.75	0.00	A
A9	4.50	0.22	21.90	81.25	6.25	A	A9	5.00	0.20	0.22	81.25	6.25	A
A12	4.00	0.44	22.84	68.75	6.25	S	A12	4.00	0.19	0.17	81.25	0.00	A
A13	4.00	0.00	14.60	86.67	0.00	A	A13	4.00	0.19	0.17	75.00	3.25	S
A15	4.00	0.25	21.88	68.75	6.25	S	A15	4.00	0.25	0.28	68.75	12.50	S
A18	4.00	0.25	16.55	73.33	0.00	S	A18	4.00	0.19	0.17	81.25	4.00	A
							A19	4.00	0.19	0.19	87.50	4.00	A
A20	4.00	0.19	16.74	81.25	0.00	A	A20	4.00	0.19	0.17	81.25	4.00	A
							A21	4.00	0.25	0.15	68.75	3.00	S
							A22	4.00	0.00	0.13	81.25	4.00	A
A24	4.00	0.19	15.98	75.00	0.00	S	A24	4.00	0.00	0.20	81.25	4.00	A

(RIR=0) + (coincidencia de opiniones = 100%) = unanimidad (U)

(RIR>0<0.25) + (coincidencia de opiniones $\geq 80\% < 100\%$) = consenso amplio (A)

(RIR $\geq 0.25 < 0.5$) + (coincidencia de opiniones $< 80\% \geq 67\%$) = consenso suficiente (S) N=16 (excepto en R1 para A18 y A13 N=15)

*Para el tratamiento estadístico se ha eliminado del análisis cuantitativo los ítems 10 y 17, por ser un ítem semicerrado con baja tasa de respuesta ($n \leq 5$), estos ítems serán objeto de análisis en la R2 con el resto de ítems de formato abierto.

Fuente. Elaboración propia.

Según los criterios establecidos, 18 ítems (casi un 80% del total), fueron consensuados en la R2, lo que representa un grado de convergencia muy alto (Sumison, 1998). A su vez, 15 de estos 18 ítems alcanzaron el consenso de al menos 3/4 de los miembros.

En la Tabla 4 se presentan los resultados de los ítems indeterminados y de los que no alcanzaron un acuerdo aceptable tanto en R1 como en R2. En la primera ronda, los resultados del análisis han mostrado que 9 ítems (37.5%) no lograron alcanzar un acuerdo mínimo. Por su parte, en la segunda vuelta los panelistas reconsideraron favorablemente 3 ítems de los 9 no consensuados en R1, concretamente, los ítems A19, A21y A22. Teniendo en cuenta que la pregunta 10b se eliminó de este análisis, sólo 5 ítems no lograron una convergencia significativa ($> 67\%$), aunque se observó que al menos el 50% de los expertos se posicionaba a favor.

La estabilidad grupal mide la respuesta conjunta del grupo cuantificando la variación de la distribución de las respuestas de una ronda en comparación con la anterior. Es una medida menos fuerte que la individual,

pero de mayor relevancia al tener en cuenta que lo que interesa es la fiabilidad de las respuestas del grupo de expertos en su conjunto (Landeta, 2002; von der Gracht, 2012).

En este sentido, se consideraría que se alcanzó un grado de estabilidad suficiente entre las Rondas 1 y 2 cuando, medida de diferentes formas, los valores fuesen los siguientes: $-0,25 \leq VRIR \leq 0,25$ y $\bullet \bullet \leq 0,27$.

Tabla 4. Resultados sin consenso de R1 y R2 (n=16)

Item	RONDA 1					RONDA 2					Grado de acuerdo
	Me	RIR	CV	% A favor [C4,C5]	% En contra [C1,C2]	Me	RIR	CV	% A favor	% En contra	
A10b	4.00	0.75	40.13	56.25	31.25						SIN CONSENSO
A11	4.00	0.75	40.13	56.25	31.25	3.50	0.86	0.49	50.00	37.50	
A14	3.00	0.33	26.36	37.50	6.25	3.50	0.29	0.26	50.00	6.25	
A16	3.00	0.67	33.97	31.25	43.75	4.00	0.25	0.20	62.50	6.25	
A19	4.00	0.25	32.36	62.50	18.75						
A21	3.50	0.29	26.22	50.00	18.75						
A22	4.00	0.44	33.54	62.50	25.00						
A23	3.50	0.29	22.85	50.00	6.25	3.50	0.29	0.26	50.00	12.50	
A25	3.00	0.58	32.55	37.50	25.00	3.50	0.29	0.27	50.00	12.50	

Me > 2 < 4 (Indeterminados)
RIR ≥ .50 (no consenso)
% de respuestas fuera del rango donde se sitúa la Me >33% (no consenso)
CV ≥ 0.26 (no consenso por exceso de variabilidad)

Fuente. Elaboración propia.

Tabla 5. Resultados de los medidores de estabilidad

Variación Coeficientes dispersión					Variación coeficientes dispersión				
Ítem	VRIR	Δ/·	·	Δ/·	Ítem	VRIR	Δ/·	·	Δ/·
A1	0.15	·	0.01	·	A14	0.05	·	0.00	·
A2	0.19	·	0.10	·	A15	0.00	·	-0.06	Δ
A3	0.20	·	0.08	·	A16	0.42	·	0.14	·
A4	0.00	·	0.03	·	A18	0.06	·	0.00	·
A5	0.14	·	0.04	·	A19	0.06	·	0.13	·
A6	0.20	·	0.06	·	A20	0.00	·	0.00	·
A7	0.00	·	0.03	·	A21	0.04	·	0.11	·
A8	0.02	·	0.03	·	A22	0.44	·	0.21	·
A9	0.02	·	0.00	·	A23	0.00	·	-0.03	Δ
A11	-0.11	Δ	-0.09	Δ	A24	0.19	·	-0.04	Δ
A12	0.25	·	0.06	·	A25	0.30	·	0.05	·
A13	-0.19	Δ	-0.03	Δ					

Estabilidad amplia : $-0,25 \leq VRIR \leq 0,25$; $\bullet \bullet > 0 < 0,20$
Estabilidad suficiente: $\bullet \geq 0,20 \leq 0,27$

Fuente. Elaboración propia.

El análisis de la variación de los dos coeficientes de dispersión (VRIR y \cdot) revela que a excepción de los ítems A16, A22 y A25, que sobrepasan los índices preestablecidos, el resto de ítems (86.96%) reflejan índices de estabilidad muy reducidos. Los signos negativos señalan un incremento en la variación, lo cual ocurre en cinco de los ítems (21.74%) y que revela que la segunda administración del cuestionario ha servido para hacer madurar las opiniones de los expertos participantes. En tal sentido, se ha observado una elevada estabilidad grupal entre las rondas.

3.2. Ítems cualitativos

Los ítems cualitativos se han dividido en dos bloques. Por un lado, se encuentran las respuestas a las cuestiones semicerradas A10a y A17a (Tabla 6), que incluían la opción "otros", y por otro lado se recogen las aportaciones realizadas por los expertos en cada una de las dimensiones en las que se dividió el cuestionario, incluyendo además el apartado final de observaciones (Tabla 7).

Tabla 6. Valoraciones de los ítems cualitativos A10a y 17a

ITEM 10a: Otros criterios no indicados en el cuestionario que debería contener la herramienta de recogida de información para ser lo más objetiva posible	Estadísticos			% acuerdo		N
	\cdot	RIR	Me	A favor - En Contra	[C4,C5] - [C1,C2]	
1º Adaptado a la población a la que se dirige (tener en cuenta en empresas donde se encuentran trabajadores inmigrantes, discapacitados, etc.)	4.47	0.25	4.00	100	0	15
2º Tener en cuenta el entorno y al trabajador como miembro de una organización, no como entes aislados	4.20	0.25	4.00	93.0	0	15
3º Ajuste al contexto de evaluación	4.15	0.13	4.00	92.0	0	13
4º La participación y la autoconfrontación	4.15	0.25	4.00	85.0	0	13
5º Las condiciones de administración	4.14	0.06	4.00	93.0	0	14
6º (Pre)conocimiento del terreno o validez ecológica	4.00	0.13	4.00	79.0	0	14
ITEM 17a: Otros sesgos de respuesta no indicados en el cuestionario que puedan atribuirse a escalas de autoinforme utilizadas en prevención de riesgos psicosociales	Estadísticos			% acuerdo		N
	\cdot	RIR	Me	A favor - En Contra	[C4,C5] - [C1,C2]	
1º En el día a día, muchas empresas siguen confundiendo conceptos tan diferentes como clima laboral o evaluación psicosocial	4.00	0.13	4.00	76.92	7.69	13
2º La mala explicación conlleva sesgos de sobreestimación	3.85	0.50	4.00	76.92	7.69	13

Me ≠ 3
Frecuencia de respuesta ≥ 75%
RIR ≤ .50

Fuente. Elaboración propia.

Respecto a las cuestiones semicerradas del ítem A10a, como recoge la Tabla 6, el grado de consenso es muy elevado. Todos los ítems presentaron una $Me \geq 4$ con porcentajes de acuerdo "favorable" que oscilan entre el 100% y el 79%. Por su parte, las cuestiones incluidas en el ítem A17a también presentaron $Me \geq 4$, si bien el porcentaje de acuerdo favorable no superó el 80%.

Finalmente, las observaciones aportadas por los expertos en la R1 e incluidas como ítems cuantitativos en la R2 también generaron un elevado consenso, tal y como muestra la Tabla 7, incluso 3 de estos 9 comentarios alcanzaron una $Me \geq 5$ y un porcentaje de acuerdo favorable del 100%.

Tabla 7. Valoraciones de los apartados I a IV y comentarios finales

	Observaciones del apartado I: "Afirmaciones generales"	Estadísticos			% acuerdo		N
		.	RIR	Me	A favor - En contra [C4,C5] - [C1,C2]		
1º	Hay una dejadez muy acusada en las empresas con respecto a las evaluaciones psicosociales porque las consideran subjetivas y poco útiles porque no se dan respuestas prácticas y rápidas a los problemas que se plantean.	4.23	0.25	4.00	92.31	0	13
2º	El hecho de tener herramientas para valorar los riesgos psicosociales no quiere decir que los valoren más que otro tipo de riesgos. Pueden tener las herramientas pero no utilizarlas o hacerlo de forma incorrecta.	4.00	0.25	4.00	76.92	00	13
	Observaciones del apartado II: "Diseño de instrumentos de medida"	Estadísticos			% acuerdo		N
		.	RIR	Me	A favor - En contra [C4,C5] - [C1,C2]		
1º	Un único instrumento de medida nos acerca al problema pero para evaluar se requieren varias herramientas complementarias. No sólo la autopercepción del trabajador, sino del grupo, además de la organización del sistema y del trabajo. Debe haber como mínimo una triangulación de datos.	4.67	0.20	5.00	100.00	0	15
2º	Algunos de los instrumentos han sido creados desde el entorno académico por lo que en muchos casos se han contentado con la obtención de datos y recomendaciones generales de actuación. Quienes venimos trabajando desde la intervención observamos que los instrumentos deben ser precisos y concretos respecto a las preguntas y respuestas. Ello facilita que en un retest puedan compararse los datos, observando y analizando las diferencias.	4.15	0.13	4.00	92.31	0	13
	Observaciones del apartado III: "Aplicación"	Estadísticos			% acuerdo		N
		.	RIR	Me	A favor - En contra [C4,C5] - [C1,C2]		
1º	Es fundamental explicar el por qué, antes de comenzar la aplicación, además de cumplir con los estándares de aplicación de cualquier tipo de test.	4.57	0.20	5.00	100.00	0	14
2º	Es importante también, la situación de la empresa en el momento en que se realiza la evaluación de riesgos (ERE, algún despido...), que la dirección esté visible durante alguna fase del proceso de la evaluación (presentación del estudio a los trabajadores, y solicitud de participación a todos...) y que quede bien garantizado el anonimato.	4.54	0.20	5.00	100.00	0	13
3º	1. Es importante, previa a la administración, que el trabajador conozca la utilidad de la cumplimentación de la batería psicosocial y para qué servirá. Es decir, que haya de antemano una buena información, y que la Dirección indique que tras el diagnóstico se intervendrá según los recursos y prioridades que se establezcan. 2. Influencia del contexto de la empresa / organización (presión agentes sociales, de la patronal, etc.).	4.53	0.20	5.00	93.33	0	15
4º	Conseguir sinceridad en las respuestas está en función, además de que esté bien hecha, de la forma de administración y de la credibilidad de la persona encargada de administrarlos.	4.36	0.22	4.50	85.71	0	14
	Observaciones del apartado de "Comentarios finales"	Estadísticos			% acuerdo		N
		.	RIR	Me	A favor - En contra [C4,C5] - [C1,C2]		
	Fundamental e imprescindible, ampliar el proceso de evaluación de factores psicosociales, ir más allá de un cuestionario, y profundizar con los trabajadores en las condiciones y entorno en el que trabajan. Llevar a cabo sesiones de trabajo en las que los trabajadores puedan participar y hacerles partícipes del proceso. La temática psicosocial debería tener un mayor peso en las organizaciones, puesto que es el origen que no se ve, de otro tipo de problemáticas organizacionales.	4.38	0.25	4.00	92.31		0

Me≠3
Frecuencia de respuesta ≥75%
RIR ≤0.5

*Sólo se muestran los ítems que cumplieron los criterios de significación descritos en la metodología.

Fuente. Elaboración propia.

4. Discusión

El presente estudio pretendía explorar las principales debilidades de evaluar riesgos psicosociales, haciendo especial hincapié en los autoinformes como herramienta de medición de las condiciones de trabajo. Para ello, a través de un estudio Delphi fueron consultados 16 reputados expertos del ámbito académico y profesional, que debían mostrar su grado de acuerdo o desacuerdo con: (a) los 25 enunciados extraídos de la revisión científica y (b) con 35 nuevos reactivos construidos a partir de la formulación individual de los propios expertos, que en su mayor parte consisten en recomendaciones para corregir los puntos débiles de las evaluaciones de riesgo psicosocial.

Los hallazgos más relevantes están en línea con los teóricos, pues los resultados revelan que los expertos están de acuerdo con casi el 80% de los ítems cuantitativos presentados. Incluso, 15 ítems alcanzaron convergencia por una amplia mayoría; es decir, al menos 3/4 de los panelistas se mostraban “de acuerdo” o “totalmente de acuerdo” con el enunciado presentado. Este hecho, aunque puede parecer intrascendente, es un modo de contrastar y validar los principales postulados extraídos de la revisión bibliográfica, de acuerdo con el sentir y la experiencia de profesionales y académicos. No es frecuente que los razonamientos derivados de la literatura científica sean refrendados posteriormente, de modo explícito y controlado, por un juicio colectivo de expertos en la materia.

Con relación al contenido de los ítems, destaca la unanimidad de consenso a favor sobre la subestimación generalizadas de los riesgos psicosociales (ítems A1, A2), tal y como se evidencia de las últimas encuestas ESENER (EU-OSHA, 2015, 2019). Este resultado cobra todavía más valor con la afirmación realizada por uno de los panelistas, quien indica que: “existe una dejadez muy acusada en las empresas con respecto a las evaluaciones psicosociales [...]”. Esta valoración fue apoyada por más del 92% de los expertos y estaría reforzando los resultados de Igartúa *et al.* (2006, citado por Cortés *et al.*, 2010) quienes, a través de un Delphi realizado a médicos del trabajo, éstos señalaron que uno de los principales obstáculos de la gestión de los riesgos psicosociales es debido a la reticencia de los directivos.

Este es un resultado inquietante ya que desde el punto de vista legal existe la misma obligación por parte de las empresas de evaluar los riesgos psicosociales que, por ejemplo, los riesgos higiénicos y mecánicos tradicionales. Si tanto la literatura revisada como el juicio experto se reafirman sobre el reiterado incumplimiento de la legislación preventiva (LPRL, 1995), debería de actuarse de inmediato sobre este foco, tanto desde el punto de vista cognitivo, aumentando la conciencia preventiva de este tipo de riesgos, como desde el punto de vista coercitivo con un mayor número de intervenciones de la ITSS.

Otro hallazgo derivado del consenso experto se refiere a que los riesgos psicosociales no sólo están subestimados, sino que también están envueltos en cierta confusión terminológica y de procedimiento (ítems A3, A4 y A11). Por ejemplo, a menudo se confunde método con técnica y se estila pensar que las evaluaciones de riesgos psicosociales consisten en aplicar un simple cuestionario (Niño, 2006). Asimismo, no es inusual encontrarse con la utilización indistinta de los términos factor psicosocial, riesgo psicosocial y factor psicosocial de riesgo, cuando realmente son nociones bien diferenciadas (Moreno y Báez, 2010). El desconcierto llega a tal punto que es habitual incluir en la evaluación de riesgos psicosociales escalas que van más allá de identificar condiciones de trabajo susceptibles de causar daño, como escalas de bienestar, de salud, clima laboral, etc., pudiendo incurrir en ese caso en intromisión profesional (Niño, 2006).

Un resultado esperado se centra en el unánime consenso a favor (100%) sobre la importancia de cumplir con las propiedades psicométricas de fiabilidad y validez (ítem 6) y de control de sesgos (ítem 7). Muy próximo se encuentra la necesidad de que las herramientas de evaluación dispongan de baremación y estandarización (ítem 8) el 93.75%; y, por último, una adecuada fundamentación teórica (ítem 9) el 81.25%. Estos resultados eran esperados porque están en consonancia con la lógica psicométrica que define la calidad de las evaluaciones de riesgos psicosociales (Abad *et al.*, 2011; Louzán, 2020).

Resulta revelador que, a pesar de que el 100% de expertos atestiguan la necesidad de controlar los sesgos, indican con un 87.50% de consenso que, según su experiencia, éstos no son controlados adecuadamente a la hora de administrar un test (ítem A19), coincidiendo con los argumentos de Dupuis *et al.* (2015). A su vez, y en consonancia con la comunidad científica, indican que los sesgos de mayor incidencia en los autoinformes son los sesgos de respuesta (ítems A12, A15, A20, A22, A21 y A24). Señalando como amenazas a la validez la falta de sinceridad, la deseabilidad social o la disimulación podrían estar distorsionando los resultados de las evaluaciones psicosociales, contaminando los resultados obtenidos y reduciendo el alcance de sus implicaciones (Louzán, 2020).

Como solución, los expertos consultados corroboran que un protocolo de administración contribuiría a minimizar dichos sesgos (ítems A18). Sin embargo, resulta paradójico que 3 de los ítems que no obtuvieron consenso tras la R2 hagan referencia explícita a los sesgos de aquiescencia (ítem A14) y autoengaño o simulación (ítem A16), y al control de los sesgos (ítem A23). Estos resultados tal vez sean debidos a la heterogeneidad de los participantes en el estudio (académicos y profesionales) y a sus intereses diferentes visiones de la evaluación de riesgos profesionales (teórica vs. pragmática).

Futuros estudios deberían reproducir estos resultados en muestras más amplias que permitan realizar análisis segmentados y comparaciones entre grupos con la debida potencia estadística.

El contrato psicológico como posible variable de distorsión es otro de los hallazgos más relevantes de este estudio Delphi. Pues un 81% de los panelistas están de acuerdo en que la percepción por parte del trabajador de incumplimiento del contrato psicológico podría influir en el falseamiento intencional de las respuestas de las escalas que miden condiciones de trabajo (ítem A24). El objetivo sería causar daño a la empresa y contrarrestar así el perjuicio padecido por las expectativas incumplidas (Guadarrama *et al.*, 2017). La escasa literatura existente acerca del impacto de la vulneración del contrato psicológico en la percepción del trabajador, hace que este resultado sea novedoso y precisamente, debido al alto porcentaje de acuerdo de los expertos consultados, sería interesante explorar empíricamente esta dimensión como posible variable de distorsión en las respuestas de los autoinformes que evalúan riesgos psicosociales.

La consulta realizada a los expertos no se agotó con los ítems cerrados, se creyó conveniente incluir preguntas abiertas con el fin de que los panelistas pudiesen expresar cuestiones no contempladas en el cuestionario o realizar sugerencias y recomendaciones prácticas para contribuir a la mejora de la calidad de las evaluaciones de riesgos psicosociales. De todas las respuestas recibidas solo se desarrollan las que cumplieron con los criterios de consenso establecidos, presentándolas con las propias palabras de los expertos para minimizar un posible sesgo de interpretación. Se desarrollan en los siguientes puntos:

a. Instrumentos adaptados a la población a la que se dirigen

En este sentido los test tienen que ser justos y no “beneficiar a ciertos grupos de población en detrimento de otros de igual nivel en el rasgo que interesa medir” (Gómez- Benito *et al.*, 2010, p. 76). Por ejemplo, un trabajador procedente de otro país puede malinterpretar los enunciados del cuestionario al atribuir un significado equivocado a las palabras. Asimismo, es necesario tener en cuenta el entorno y al trabajador como miembro de una organización, no como entes aislados. En la misma línea, se apunta la necesidad de ajustar el cuestionario al contexto de evaluación. Conviene no utilizar escalas genéricas para toda la organización, pues cada puesto contiene una realidad psicosocial diferente, y sería preciso evaluar por puestos de trabajo o por colectivos de trabajadores (Nogareda y Almodóvar, 2006).

b. La participación del trabajador es fundamental

La participación de los trabajadores en las actividades de prevención, incluida la evaluación, es otro aspecto que destacan como crucial. Además, es una obligación establecida por Ley (arts.18.2, 34 y 36.2, LPRL, 1995). De acuerdo con los expertos consultados, se deben confrontar los resultados con los trabajadores y llevar a cabo sesiones de trabajo en las que los trabajadores puedan participar y hacerles partícipes del proceso, en línea con las indicaciones del reciente Criterio Técnico de la ITSS (2021). Sin embargo, no existen directrices en forma de manuales o guías de recomendación sobre cuál es el mejor modo de participación, y menos de cómo abordar la confrontación de los resultados, lo que puede contribuir a evaluaciones menos precisas.

c. Triangulación de datos

El 100% de los miembros del panel confirman que para refrendar la percepción individual del trabajador es necesario utilizar la triangulación de datos. Precisamente, la triangulación metodológica es una de las principales recomendaciones para obtener evaluaciones precisas y certeras. Consiste en la combinación de dos o más metodologías, generalmente cualitativa y cuantitativa, para obtener una aproximación más real del fenómeno objeto de estudio. Es una de las medidas más apreciadas en la literatura para dotar de mayor

calidad a la evaluación de riesgos; concretamente se recomienda utilizar las entrevistas como complemento de los cuestionarios de autoinforme (ITSS, 2012, 2021; Moreno y Báez, 2010; Oncins y Almodóvar, 1997).

d. Incrementar la validez aparente de los instrumentos

Casi un 93% de panelistas han confirmado estar de acuerdo con el ítem que indica que con frecuencia las herramientas de prevención creadas desde el entorno académico pecan de formular ítems y medidas correctoras demasiado generales. Debido a la distinta procedencia profesional de los expertos consultados sería interesante reflexionar sobre ello e incrementar la validez aparente de los instrumentos colaborando, para ello, con los profesionales de la PRL. Una posible solución para conseguir instrumentos más precisos en cuanto a preguntas y respuestas, tal y como reclaman desde la rama intervencionista, sería la cooperación y colaboración de todos los sectores implicados y preocupados por la prevención.

e. Buenas prácticas en el uso de los test. Recomendaciones de los panelistas

Los expertos indican que deben cuidarse las condiciones de administración de los cuestionarios ya que existen variedad de factores que pueden contribuir a distorsionar los resultados obtenidos y por ende las generalizaciones efectuadas. Entre algunas de sus recomendaciones se encuentran:

- e.1. Explicar el objetivo y la utilidad de los cuestionarios antes de repartirlos;
- e.2. Implicar a la dirección durante al menos alguna fase del proceso de evaluación;
- e.3. Tener en cuenta la situación de la organización en el momento de evaluar, por ejemplo, si la empresa está incurso en algún ERE o se esperan despidos los trabajadores estarán menos receptivos, y por lo tanto sus respuestas pueden llegar a estar sesgadas;
- e.4. Asegurar la confidencialidad y el anonimato de la información obtenida;
- e.5. Garantizar la credibilidad de las personas que lleven a cabo la evaluación de riesgos psicosociales;
- e.6. Tener en cuenta factores externos a la hora de administrar los test, como las condiciones térmicas inadecuadas, el ruido, la ansiedad de los trabajadores al ser encuestados o una preparación inadecuada del administrador de los test.

En síntesis, el presente estudio Delphi ha permitido obtener información valiosa sobre las principales debilidades de la evaluación de los riesgos psicosociales desde el juicio experto. En líneas generales, los resultados muestran un elevado grado de consenso. Por una parte, en la mayoría de los ítems cerrados los expertos concuerdan en los desafíos conceptuales y aplicados que encarna la evaluación de los riesgos psicosociales y, por otra parte, la riqueza de aportaciones realizadas por los expertos en los ítems abiertos, ha contribuido a ampliar considerablemente la calidad del estudio, pues constituyen una fuente valiosa de información basada en la experiencia y conocimiento experto.

Concluyendo, se han identificado directrices prácticas que podrían contribuir a superar las amenazas identificadas, y mejorar así la utilidad práctica del presente trabajo. Sobre todo en estos tiempos, donde la reciente pandemia mundial ha empeorado el panorama psicosocial laboral y, más que nunca, hace falta realizar evaluaciones de riesgo psicosocial de calidad. Evaluaciones con capacidad de reflejar resultados lo más cercanos posibles a la realidad, esto es, con el menor error posible.

A pesar del rigor con el que se ha diseñado y llevado a cabo el presente estudio, éste no carece de limitaciones. Lo que primero podría ser objeto de crítica es la consabida subjetividad en la que está basada la metodología, lo que llevaría a cuestionar el grado de conocimiento real de los expertos en el panel Delphi. Esta debilidad dependerá del acierto y destreza en la elección de panelistas. Para salvar esta limitación se ha tratado con especial esmero el proceso de selección de participantes, aplicando criterios de inclusión rigurosos como el coeficiente K.

En lo que respecta al tamaño de la muestra (16 panelistas), aunque no limita de ningún modo los resultados obtenidos por estar dentro del rango recomendado, no permite aprovechar toda la información proporcionada por los expertos y explorar otras cuestiones ajenas al consenso en sí. Por ejemplo, una muestra mayor hubiera permitido comprobar la existencia de subgrupos, así como examinar si las divergencias aparecidas han sido fruto de la diferencia de intereses en la población seleccionada. En consecuencia, sería de interés que futuras investigaciones pudiesen replicar este estudio en una muestra más amplia que permita la detección de subgrupos.

Asimismo, las cuestiones indeterminadas en la que no se consiguió un consenso experto (A11, A14, A15, A16, A23 y A25), dada su irresolución desde la esfera cualitativa, deberían ser objeto de una exploración más profunda que permita comprobar de modo empírico su pertinencia, como por ejemplo explorar el impacto de la DS en los autoinformes que evalúan riesgos psicosociales.

La consulta a los panelista se realizó antes de la irrupción de la inesperada pandemia mundial por el virus COVID-19. Desde entonces, el escenario ha empeorado drásticamente (Tsamakis *et al.*, 2020). En lo referente al ámbito laboral, el Covid-19 ha incrementado la exposición adversa a los factores de riesgo psicosocial, como el miedo a perder el trabajo, recortes, despidos, reducción de beneficios, entre otros (OIT, 2020). Sería interesante realizar un estudio similar teniendo en cuenta el escenario post pandemia. A priori, se presume que no deberían de encontrarse resultados muy alejados, pues aunque haya aumentado la exposición adversa a muchos estresores, el modo de evaluar los riesgos psicosociales y las herramientas de medición no han variado significativamente.

Referencias

- Abad, F.J., Olea, J., Ponsoda, V. y García, C. (2011). *Medición en ciencias sociales y de la salud*. Síntesis.
- Baratech, M. (2006). El estrés: epidemia laboral del siglo XXI. *Fomento de la Producción*, 1262, 10-11.
- Boix, P., Gil, J.M. y Rodrigo, F. (2008). Prioridades estratégicas para la mejora del sistema de prevención de riesgos laborales en España: un estudio Delphi entre profesionales de la salud y seguridad en el trabajo. *Archivos de Prevención de Riesgos Laborales*, 11(1), 20-26.
- Brill, J., Bishop, M. & Walker, A. (2006). An investigation into the competencies required of an effective project manager: A Web-based Delphi study. *Educational Technology Research & Development*, 54(2), 115-140.
- Burguet-Lago, I., Rodríguez-Rabelo, A., y Jorge-Chacón, D. (2019). Aplicación de tecnologías para la determinación de la competencia de expertos. *Revista cubana de Ciencias Informáticas*, 13(1), 116-126.
- Cabero, J. y Llorente, M. C. (2013). La aplicación del juicio de experto como técnica de evaluación de las tecnologías de la información (TIC). *Revista de Tecnología de Información y Comunicación en Educación*, 7(2)11-22. <http://servicio.bc.uc.edu.ve/educacion/eduweb/v7n2/art01.pdf>
- Cortés, I., Artazcoz, L., Igartua, A., Juanola, E., Ladona, C., Llonch, A.,... Molinero, E. (2010). *Guía de buena praxis para la vigilancia de la salud mental relacionada con los factores de riesgo psicosocial*. Generalitat de Catalunya. Departament de Salut. Direcció General de Salut Pública.
- Dalkey, N.C. (1972). The Delphi method: an experimental application of group opinion. In N.C. Dalkey, D.L. Rourke, R. Lewis & Snyder D. (Eds.), *Studies in the quality of life* (pp. 13-54). Lexington Books.
- Dalkey, N.C., & Helmer, O. (1963). An experimental application of the delphi method to the use of experts. *Management Science*, 9(3), 458-467.
- Dupuis, M., Meier, E., Capel, R. & Gendre, F. (2015). Measuring individuals' response quality in self-administered psychological tests: an introduction to Gendre's functional method. *Frontiers in Psychology*, 6(629), 1-12. <https://doi.org/10.3389/fpsyg.2015.00629>
- EU-OSHA (2015). *ESENER-2: Segunda Encuesta europea en las empresas sobre riesgos nuevos y emergentes. Primeros resultados*. EU-OSHA. <https://bit.ly/3zGVSMa>
- EU-OSHA (2019). *Third European Survey of Enterprises on New and Emerging Risks (ESENER 3)*. EU-OSHA. <https://bit.ly/3zAniDz>
- Ferrán, M. (2001). *SPSS para Windows. Análisis estadístico*. Osborne McGraw-Hill.
- Ferrer, R., Guilera, G. y Peró, M. (2011). *Propiedades Psicométricas del Instrumento de Valoración de Riesgos Psicosociales del Instituto Nacional de Seguridad e Higiene en el Trabajo (FPSICO)*. Universitat de Barcelona.
- Gil-Monte, P.R. (2010). Situación actual y perspectiva de futuro en el estudio del estrés laboral: la Psicología de la Salud Ocupacional. *Información psicológica*, 100, 68-83.
- Gómez-Benito, J., Hidalgo, M.D. y Guilera, G. (2010). El sesgo de los instrumentos de medición. *Tests justos. Papeles del Psicólogo*, 31(1), 75-84.
- Guadarrama, A.I.M., Arroyo, J.C., y Flores, G.R. (2017). Efecto de la violación del contrato psicológico y el agotamiento emocional sobre el cinismo del empleado. *Estudios Gerenciales*, 33(143), 124-131. <https://doi.org/10.1016/j.estger.2017.04.002>

- Hernández-Sampieri, R. y Mendoza, C (2018). *Metodología de la investigación. Las rutas cuantitativa, cualitativa y mixta*. McGraw Hill Education.
- Inspección de Trabajo y Seguridad Social (ITSS) (2012). *Guía de Actuaciones de la Inspección de Trabajo y Seguridad Social sobre Riesgos Psicosociales*. ITSS. <https://bit.ly/3p1MNsU>
- Inspección de Trabajo y Seguridad Social (2021). *Criterio Técnico OE ITSS nº 104/2021 sobre actuaciones de la Inspección de Trabajo y S.S. en Riesgos Psicosociales*. ITSS. <https://bit.ly/3p1CSTD>
- Instituto Nacional de Seguridad y Salud en el Trabajo [INSST] (2018). *El efecto sobre la salud de los riesgos psicosociales en el trabajo: una visión general*. INSHT. <https://apps.who.int/iris/handle/10665/310949>
- Instituto Nacional de la Seguridad Social [INSST] (2022). *Directrices básicas para la gestión de los riesgos psicosociales*. INSST. <https://bit.ly/3PkkBap>
- Landeta, J. (2002). *El método Delphi. Una técnica de previsión del futuro*. Ariel.
- Ley 31/1995, de 8 de noviembre, de Prevención de Riesgos Laborales. *Boletín Oficial del Estado*, 10 de noviembre de 1995, núm 269, pp. 32590-32611.
- López-Gómez, E. (2018). El Método Delphi en la Investigación Actual en Educación: Una Revisión Teórica y Metodológica. *Educación XXI*, 21(1), 17-40. <https://doi.org/10.5944/educxx1.20169>
- Louzán, R. (2020). Mejorar la calidad de las evaluaciones de riesgos psicosociales mediante el control de sesgos. *Archivos de Prevención de Riesgos Laborales*, 23(1), 68-81. <https://dx.doi.org/10.12961/apr.2020.23.01.06>.
- Marín-González, F., Pérez-González, J., Senior-Naveda, A. y García-Gulinay, J. (2021). Validación del diseño de una red de cooperación científico-tecnológica utilizando el coeficiente K para la selección de expertos. *Información tecnológica*, 32, 79-88. <http://dx.doi.org/10.4067/S0718-07642021000200079>
- Moncada, S., Llorens, C., Salas, S., Moriña, D. y Navarro, A. (2021). La tercera versión de COPSOQ-ISTAS21. Un instrumento internacional actualizado para la prevención de riesgos psicosociales en el trabajo. *Revista Española de Salud Pública*, 95 .
- Moreno, B., y Báez, C. (2010). *Factores y riesgos psicosociales, formas, consecuencias, medidas y buenas prácticas*. Instituto Nacional de Seguridad e Higiene en el Trabajo. Universidad autónoma de Madrid.
- Niño, J. (2006). Los errores en las evaluaciones de los riesgos psicosociales: prevenir la confusión. *MAPFRE Seguridad* (103), 25-35
- Nogareda, C., y Almodóvar, A. (2006). *El proceso de evaluación de factores psicosociales*. NTP 702. (INSHT).
- Okoli, C. & Pawlowski, S. (2004). The Delphi method as a research tool: an example, design considerations and applications. *Information & Management*, 42,15-29. <https://doi.org/10.1016/j.im.2003.11.002>
- Oncins, M. y Almodóvar, A. (1997). *Factores Psicosociales: fases para su evaluación (NTP 450)*. INSHT.
- Organización Internacional del Trabajo (O.I.T.) (2020). *Managing work-related psychosocial risks during the COVID-19 pandemic*. ILO. <https://www.ilo.org/global/topics/coronavirus/lang--es/index.htm>
- Real Decreto 39/1997, de 17 de enero, por el que se aprueba el Reglamento de los Servicios de Prevención. *Boletín Oficial del Estado*, 31 de enero de 1997, núm. 27, pp. 3031-3045.
- Steurer, J. (2011). The Delphi method: an efficient procedure to generate knowledge. *Skeletal Radiol*, 40, 959961.
- Sumison, T. (1998). The Delphi Technique: An Adaptive Research Tool. *British Journal of Occupational Theory*, 67(4), 153-156. <https://doi.org/10.1177/030802269806100403>
- Tsamakis K., Triantafyllis A.S., Tsiptsios D., Spartalís E., Mueller C., Tsamakis C. & Rizos E. (2020). COVID-19 related stress exacerbates common physical and mental pathologies and affects treatment (Review). *Experimental and Therapeutic Medicine*, 20(1), 159-162. <https://doi.org/10.3892/etm.2020.8671>
- Velásquez, M. (2020). Cómo se está legislando sobre los riesgos psicosociales en el trabajo en la Unión Europea. En Correa Carrasco, M., Quintero Lima, M. G. (Coord.). *Los nuevos retos del trabajo decente: la salud mental y los riesgos psicosociales* (Objetivos de Desarrollo Sostenible: 3, 5, 8, 10), (pp. 38-62). Universidad Carlos III de Madrid.
- Von der Gracht, H.A. (2012). Consensus measurement in Delphi studies: Review and implications for future quality assurance. *Technological Forecasting and social change*, 79(8), 1525–1536.

Breve CV de la autora

Rita Louzán Mariño es doctora en Procesos Psicológicos y Comportamiento Social por la USC. Profesora en la Escuela Superior de Ingeniería y Tecnología de la UNIR. Miembro desde el año 2017 del Comité Técnico de Normalización en AENOR sobre Riesgos Psicosociales. Sus líneas de investigación se centran en la mejora de los métodos de prevención de los riesgos psicosociales y ergonómicos de origen laboral.

Agradecimientos

La autora del presente trabajo quiere manifestar su más sincero y profundo agradecimiento a los 16 panelistas expertos encuestados por su desinteresada participación en el Estudio Delphi, a saber: Josep M. Blanch Ribas, Eduardo García Cueto, Pedro R. Gil Monte, Joan Guardiola Olmos, Daniel Iglesias Pastrana, María José Lasa Gorraiz, Francisco López Barón, Javier Llana Álvarez, Jenny M^a Manteiga Rilo, Albert Mariné, Margarita Oncins de Frutos, Gabriel Pérez Zambrana, Mari Carmen Ramos Lema, Marisa Salanova Soria, Guillermo Soriano Tarín, y Placer Vieco Marcos.