



**Universitat de les  
Illes Balears**

*Títol: ¿Qué tipo de instrumentos de medición del estrés psicológico se aplican en profesionales de enfermería y cuál es su validación psicométrica?*

---

NOM AUTOR: *Marta León González*

---

DNI AUTOR: *43189069-Z*

---

NOM TUTOR: *Joana Fornés Vives*

---

## **Memòria del Treball de Final de Grau**

Estudis de Grau de *ENFERMERÍA*

---

Paraules clau: *Enfermería, Estrés psicológico, Cuestionarios, Estudios de validación, Psicometria*

---

de la

UNIVERSITAT DE LES ILLES BALEARS

Curs Acadèmic: *2013/2014*

---

Cas de no autoritzar l'accés públic al TFG, marqui la següent casella:

## **Resumen**

**Introducción:** El estrés es un concepto de múltiples perspectivas que ha sido estudiado desde marcos teóricos muy concretos y de tipo ocupacional. Por otro lado, el profesional de enfermería es uno de los más afectados a nivel laboral y personal por el estrés psicológico. Este hecho se relaciona con la complejidad y alto componente humano de su profesión. Esta heterogeneidad de ambos elementos demanda de escalas psicométricas validadas y fiables para medir la naturaleza del estrés en el entorno enfermero.

**Objetivos:** explorar los instrumentos de medición del estrés en personal de enfermería y sus validaciones psicométricas.

**Metodología:** búsqueda bibliográfica de la literatura científica de los últimos 20 años con criterios de inclusión y exclusión específicos en PubMed/Medline, EBSCOhost y ScienceDirect siguiendo las recomendaciones del estándar PRISMA.

**Resultados:** Los instrumentos más usados en la medición del estrés fueron el “Effort–Reward Imbalance Questionnaire” (ERI), el “Job Content Questionnaire” (JCQ) y la “Nursing stress scale” (NSS) y estuvieron dirigidos a enfermeras generalistas de grandes hospitales en más del 70% de los estudios. Las validaciones más aplicadas fueron la convergente-discriminante y la consistencia interna por Alfa de Cronbach.

**Discusión:** se describió un alto componente de variabilidad en los instrumentos encontrados con escasa aplicabilidad de aquellos con más calidad psicométrica a muestras de enfermeras especializadas. La validez de constructo a pesar de ser un elemento psicométrico imprescindible, fue de escaso rigor y aplicación.

**Conclusiones:** son necesarios más estudios de instrumentos para enfermeras especializadas y validaciones psicométricas rigurosas para los explorados en esta revisión.

**Palabras clave:** *Enfermería, Estrés psicológico, Cuestionarios, Estudios de validación, Psicometria*

## Introducción

El estrés es un concepto multidimensional que puede ser abordado desde distintos marcos teóricos dependiendo de la perspectiva de su estudio<sup>1</sup>. Sólo en el ámbito psicosocial, puede significar conceptos tan distintos como; ansiedad, esfuerzo, tensión emocional, sobrecarga, fatiga o frustración<sup>2</sup>. Algunos investigadores conciben el estrés como una respuesta física, conductual o emocional. Para otros se trata de un estímulo ambiental derivado de situaciones específicas (estresores) mientras que para unos terceros, este constructo es una interacción entre el ambiente externo y el sujeto<sup>3,4</sup>.

Diversos autores han descrito sus propios modelos o marcos teóricos sobre el estrés a nivel psicosocial y laboral. Los más valorados son el modelo transaccional de Lazarus y Folkman (1984)<sup>5</sup>, el modelo de demanda-control (y apoyo social) de Karasek y Theorell (1990)<sup>6</sup> y el modelo de desequilibrio de esfuerzo-recompensa de Siegrist (1996)<sup>7</sup>. El primero describe el estrés como una respuesta a un desequilibrio entre las demandas o exigencias del ambiente y los recursos para responder de manera efectiva. En el segundo el estrés nacería de la interacción del sujeto con unas variables psicosociales específicas (estresores): altas demandas psicológicas, bajo control en la toma de decisiones y bajo apoyo social. En el tercero y último el estrés se generaría a partir de un elevado esfuerzo, un salario inadecuado y un bajo control sobre el propio estatus ocupacional.

Siguiendo esta línea, uno de los sectores laborales más afectados por el estrés es el sanitario, especialmente el área de enfermería dadas las características laborales de la profesión. La enfermera vive de cerca la enfermedad, la invalidez, las necesidades del individuo, la muerte, el sufrimiento así como los dilemas éticos, demandas y conflictos profesionales derivados de su bien intrínseco: “cuidar”<sup>8-11</sup>. A esto se le suma que, todavía hoy en día la enfermería es un colectivo mayoritariamente femenino que ha de hacer frente a tal sobrecarga de tareas laborales y familiares que corre un elevado riesgo de desgaste emocional y de deterioro personal<sup>12,13</sup>. Por lo tanto, el estrés de la enfermería no procede de una única causa ni de varios estresores aislados; su etiología es el resultado de un cúmulo de factores estresantes asociados en el desempeño de la profesión y que sobrepasan los recursos de afrontamiento personales. Se puede considerar, pues, que el estrés es inherente del desarrollo de la actividad enfermera y sus consecuencias derivan en pérdida de motivación, descenso del rendimiento, deterioro de la calidad del cuidado, insatisfacción laboral, problemas de salud e incluso abandono profesional<sup>14,15</sup>.

Así pues, tanto la variabilidad del constructo del estrés como la heterogeneidad y complejidad del puesto de trabajo de enfermería, generan dificultades en su estudio, especialmente en lo relacionado a su medición. La medición consiste en reglas para la asignación de números a objetos en tal forma que representen cantidades de atributos (objetos o personas) <sup>16</sup>. Sin embargo, en el caso del estrés se trata de medir algo no físico ni directamente observable. No existe ningún parámetro simple que refleje, de forma fiable, el grado de estrés individual, y es necesario recurrir a un conjunto de pruebas que registren los distintos índices individuales de respuesta. Por ello se han utilizado diversas escalas psicométricas, es decir, procedimientos estandarizados compuestos por ítems seleccionados y organizados, concebidos para provocar en el individuo ciertas reacciones registrables. Estas reacciones pueden ser de toda naturaleza en cuanto a su complejidad, duración, forma, expresión y significado. Sin embargo, la legitimidad y eficiencia de estos tests psicométricos depende de una serie de cualidades como la fiabilidad y la validez. La fiabilidad (o consistencia) de un test es la precisión con que el test mide lo que mide, en una población determinada y en las condiciones normales de aplicación. En otras palabras, es el grado de precisión del test. La validez es el grado en que cada test refleja el constructo que dice medir y que las relaciones entre tests que miden distintos constructos reflejan las relaciones hipotetizadas entre ellos <sup>16-20</sup>.

## **Objetivos**

- Explorar los instrumentos de medición de estrés psicológico en el contexto de enfermería.
- Describir el tipo de validación psicométrica usada incidiendo en su calidad metodológica.

## **Estrategia de búsqueda bibliográfica**

Revisión sistemática de la literatura científica de los últimos 20 años dirigida a cumplir los objetivos descritos sobre instrumentos de medición de estrés en personal de enfermería y su validación psicométrica. Esta metodología está guiada por las premisas y recomendaciones del estándar de PRISMA<sup>21</sup> para asegurar un proceso de recuperación de la información y la reproducción de las búsquedas realizadas.

Criterios de inclusión (CI):

- CI-1 Artículos en inglés, castellano y catalán.
- CI-2 Artículos cuyo contenido esté centrado en el estudio de las características psicométricas de un instrumento, escala o test del estrés psicológico en personal de enfermería publicados entre 1994 y 2014.
- CI-4 Estudios que midieran el estrés no sólo a través de un único estímulo (conflicto con el médico, muerte del paciente, cargas de trabajo...) u efecto (burnout, insomnio, dolor, abandono de la profesión...) sino a través de varios tipos de estresores o consecuencias.

Criterios de exclusión (CE). Debido a la alta complejidad y heterogeneidad de algunos factores asociados al estrés psicológico se excluyen:

CE-1 Estudios centrados en patologías de salud mental como la depresión o estrés postraumático.

CE-2 Estudios sobre alumnos de enfermería o recién titulados dado que las características del estrés pueden variar respecto a las expresadas por personal laboral.

CE-3 Estudios centrados en intervenciones de prevención, afrontamiento o tratamiento del estrés.

#### Búsqueda bibliográfica.

Las palabras clave y descriptores para generar la búsqueda bibliográfica fueron extraídos del MeSH - Medical Subject Headings de la U.S. National Library of Medicine (NLM), del vocabulario estructurado y trilingüe DeCS - Descriptores en Ciencias de la Salud y de palabras clave de artículos publicados en revistas indexadas en la Journal Citation Reports (JCR). Los términos se presentan en la tabla 1 en los dos idiomas principales de búsqueda:

Se realizaron búsquedas con aproximadamente 20 combinaciones de estos descriptores y booleanos en las siguientes bases de datos: PubMed/Medline, Academic Search Premier, PsycINFO, CINAHL with Full Text, Abstracts in Social Gerontology, Violence & Abuse Abstracts, Family Studies Abstracts y PsycARTICLES (todos ellos a través de EBSCOhost Research Databases). Por último también se usó el recurso ScienceDirect – Sciverse como método de acceso a revistas publicadas por Elsevier, Academic Press, North Holland, Pergamon, Harcourt Health Sciences entre otros editores.

*Tabla 1: Principales descriptores de búsqueda.*

	CASTELLANO	INGLÉS
Raíz 1	Estrés psicológico	“Psychological stress”
Raíz 3	Enfermería	“Nursing”
Raíz 2	Cuestionario	“Questionnaire”
Secundario 1	Enfermera	“Nurse”
Secundario 2	Estrés	“Stress”
Secundario 3	Estrés laboral	“Job stress”
Secundario 4	Escala	“Scale”
Secundario 5	Estudios de validación	“Validation studies”
Secundario 6	Medición	“Measurement”
Secundario 7	Desarrollo de instrumento	“Instrument development”
Marginal	Personal enfermería	“Nursing staff”
Marginal	Psicología	“Psychology”
Marginal	Psicometría	“Psychometrics”

#### Algoritmos de búsqueda

De las posibles combinaciones de booleanos y descriptores sólo 10 algoritmos ofrecieron resultados de artículos óptimos y acordes a los objetivos de búsqueda.

Para PubMed/Medline:

- “Nursing” AND “Stress, Psychological” AND “Questionnaire”
- "Nursing" AND "Stress" AND "Psychometrics" AND "Questionnaire".
- "Nursing Staff" AND “Psychology" AND "Questionnaires” AND "Stress, psychological".
- "Nursing" AND “Stress AND “Scale” AND “Psychometric”.
- “Stress” AND “Validation studies” AND “Nursing”.

Para EBCOhost:

- “Nursing” AND “Stress” AND “Psychometric”
- “Nurse” AND “Stress psychology” AND “Psychometric”
- “Psychometric” AND “Job stress” AND “Nursing”
- “Instrument development” AND “Job stress” AND “Nursing”

Para ScienceDirect

- “Nursing” AND “Stress” AND “Psychometric”
- “Nurses” AND “Stress” AND “Questionnaire”

#### Análisis y elección de los estudios: escalones de selección y revisión

La revisión de la literatura hallada se dividió en tres pasos de análisis dependiendo de la profundidad de la lectura de los estudios y su posterior selección: un primer paso en el que se estudiaban el título y resumen de los artículos; un segundo paso en el que se analizaba el texto completo de forma superficial buscando el cumplimiento del objetivo general de estudio; y un tercer paso en el que se profundizaba y concretaba sobre la calidad de los estudios y objetivos secundarios.

Dadas las características de los principios de inclusión y exclusión de esta revisión, el CI-1 fue aplicado a través del primer paso de revisión. Para los criterios de exclusión CE-1, CE-2 y CE-3 fueron necesarios tanto el primer como el segundo paso de análisis, mientras que para los criterios de inclusión 2 y 3 se siguieron los dos últimos pasos.

Todo el proceso de inclusión y exclusión de artículos y los resultados finales están representados en un diagrama de flujo (Figura 1) que sigue el orden establecido por escalones de revisión descritos anteriormente.

### **Resultados**

En las búsquedas se obtuvieron un total de 1138 artículos. Tras la aplicación de los criterios de inclusión y exclusión, sólo 21 artículos<sup>22-42</sup> fueron seleccionados para la revisión (figura 1): 11 procedentes de la base de datos PubMed/Medline<sup>22-32</sup>; 5 encontrados en distintas revistas a través del servicio ScienceDirect-Sciverse<sup>33-37</sup>. Otros 3 artículos se extrajeron del metabuscador EBCOhost, excluyendo repeticiones de las anteriores búsquedas<sup>38-40</sup>. El resto se extrajeron como bibliografía secundaria de los artículos anteriormente encontrados<sup>41,42</sup>.

En el apartado final de esta revisión se puede encontrar una tabla resumen (Anexo 1) de los contenidos principales de los artículos hallados en esta revisión. Está dividida en 5 apartados que facilitan la lectura: autor y año, nombre del instrumento, muestras de estudio, características cualitativas de la escala o test y tipo de validación psicométrica.

#### Tipo de estudios y objetivos

Aproximadamente el 58% de los artículos fueron estudios descriptivos y de validación psicométrica de instrumentos originales, inicialmente creados para el estrés ocupacional y aplicados o adaptados a muestras de enfermería diversas<sup>23,28-30,33,34,38-42</sup>. El resto procedieron de estudios de creación y desarrollo de instrumentos originales únicamente creados para personal de enfermería<sup>22-25-27,31,32,35-37</sup>.

### Tipo de instrumentos y marco conceptual

Los instrumentos más usados en la medición del estrés en personal de enfermería fueron el “Effort–Reward Imbalance Questionnaire” (ERI) que aparece en 5 estudios<sup>23,29,38,40,42</sup>; el “Job Content Questionnaire” (JCQ) que fue aplicado en 2 artículos<sup>28,33</sup>; la “Nursing stress scale” (NSS) descrita en dos ocasiones<sup>26,27</sup> y “The Stress of Conscience Questionnaire” (SCQ) que se muestra en dos estudios<sup>30,41</sup>. Todos estos instrumentos aparecen en, aproximadamente, el 50% de los estudios encontrados en la revisión y representan, mayoritariamente, escalas originales creadas para estrés ocupacional general pero adaptadas a personal de enfermería. La otra mitad de estudios corresponden a escalas de una única aplicación por estudio entre las que se encuentran las siguientes: “The charge nurse stress questionnaire”(CNSQ)<sup>22</sup>, Work Stress Inventory for nurses in Oncology”(WSINO)<sup>24</sup>, “The Nurse Stress Checklist”(NSC)<sup>37</sup>, “Medical Personnel Stress Survey” (MPSS-R)<sup>34</sup>, “Post-Code Stress Scale”<sup>25</sup>, “ Tool to assess levels of stress and burnout”<sup>31</sup>, The Psychiatric Nurse Job Stressor Scale” (PNJSS)<sup>32</sup>, “Stressor Scale for Clinical Research Coordinators” (SSCRC)<sup>35</sup>, “Demand-control.support questionnaire”(DCS)<sup>38</sup>, “New Job-demand scale”<sup>36</sup> y finalmente Strain in Dementia Care Scale (SDCS)<sup>30</sup>.

Por otra parte, el marco conceptual del estrés en personal de enfermería más aplicado fue el de “Desequilibrio Esfuerzo-Recompensa”<sup>23,29,38,40,42</sup> creado por Siegrist<sup>43-45</sup> y expresado en el cuestionario ERI. Le sigue el modelo teórico de demanda-control<sup>23,28,33,36</sup> de Karasek y Theorell expuesto en los tests “Job Content Questionnaire”, “Demand-control.support questionnaire”(DCS) y “New Job-demand scale”<sup>46-48</sup>. Otros marcos teóricos menos usados fueron el de la teoría del rol en las expectativas, identidades y comportamientos de Biddle<sup>49</sup> (Post-Code Stress Scale) y el modelo transaccional de Lazarus y Folkman<sup>4</sup> (The Nurse Stress Checklist”). Fuera del marco teórico, la perspectiva del estrés más aplicada fue la de estrés como estímulo (estresor) que se puede encontrar en instrumentos como “The charge nurse stress questionnaire”, “Work Stress Inventory for nurses in Oncology”, “The Psychiatric

Nurse Job Stressor Scale”, “Strain in Dementia Care Scale” y “Stressor Scale for Clinical Research Coordinators” o la “Nursing Stress Scale”. Otras escalas determinaron el estrés desde la perspectiva de respuesta como la “Medical Personnel Stress Survey” que medía alteraciones somáticas relacionadas con sentimientos asociados a varias situaciones de enfermería. Por último, la “Stress of Conscience Questionnaire” estudió la interacción entre ambiente y sujeto con sus características personales y los conflictos socio-laborales. Contrariamente, algunas escalas no realizaron una adecuada contextualización o conceptualización del estrés, como es el caso de la “Tool for burnout and stress”.

#### Instrumentos de medida y muestras de estudio.

Los instrumentos fueron adaptados a enfermeras de varios países a nivel mundial. De Europa se incluyeron estudios en España<sup>26,33,39,42</sup>, Dinamarca<sup>40</sup>, Suecia<sup>41,30</sup>, Francia<sup>24</sup> y Grecia<sup>29</sup>. En América fueron de Méjico<sup>39</sup>, Colombia<sup>39</sup>, Brasil<sup>32</sup> y Estados Unidos<sup>25</sup> mientras que en Asia y Oceanía encontramos enfermeras de China<sup>27,28,34,37</sup>, Israel<sup>22</sup>, Japón<sup>32,35</sup> y Australia<sup>31</sup>.

Aproximadamente un 72% de los instrumentos validados fueron aplicados en el ambiente hospitalario<sup>22-29,33,35,36,38,41,42</sup>. De estos, un 19% se administraron en unidades específicas de hospitalización como, oncología<sup>24</sup>, cuidados intensivos y urgencias<sup>25,27,34</sup>. Por otro lado, aproximadamente un 19% fueron instrumentos destinados a niveles extra-hospitalarios como residencias de ancianos<sup>30</sup>, centros de psiquiatría<sup>32</sup> o atención primaria<sup>36,40</sup>.

#### Validaciones psicométricas y calidad metodológica

Para describir los resultados de validación psicométrica, esta revisión siguió algunos criterios básicos sugeridos por autores como Cronbach<sup>50</sup> Norbeck<sup>51</sup> o Cortada de Kohan<sup>20</sup> que establecen la existencia de cuatro estándares mínimos necesarios de validación de un instrumento: validez de constructo, validez convergente-discriminante, consistencia interna y fiabilidad por test-retest. Para apoyar estos datos se realizó un estudio de la calidad metodológica de los instrumentos hallados basada en las normas de Carretero-Dios<sup>45</sup>. Esta serie de recomendaciones para el desarrollo y revisión de estudios instrumentales estuvo representada por 7 fases:

1. Justificación del estudio.
2. Delimitación conceptual del constructo a evaluar.
3. Construcción y evaluación cualitativa de ítems.
4. Análisis estadístico de los ítems.

5. Estudio de la dimensionalidad del instrumento (estructura interna).
6. Estimación de la fiabilidad.
7. Obtención de evidencias externas validez.

La descripción completa de sus características se puede encontrar en el Anexo 2.

Las fases 5,6 y 7 hablan de la aplicación de validaciones psicométricas a un nivel más riguroso pero incluyen la existencia de las más básicas (validez de constructo, validez convergente-discriminante, fiabilidad y consistencia interna).

Para una mejor descripción de los resultados, se decidió crear dos tablas (2 y 3) con tres grados de cumplimiento de las siete normas o fases:

- Cumplimiento positivo (si): Cumple la fase y responde a todas las preguntas establecidas para ésta:
- Cumplimiento incompleto (dudoso o incompleto): Responde únicamente a la mitad de las preguntas establecidas o no cumple con todos los criterios descritos. Algunos ejemplos se muestran en el anexo 2.
- Cumplimiento negativo (NO): No cumple con ninguno de los apartados de cada fase.

De los resultados obtenidos para tipos de validación psicométrica y análisis metodológico encontramos los siguientes datos:

De los 21 artículos encontrados sólo dos <sup>29, 36</sup> cumplieron con las siete recomendaciones establecidas por fases de Carretero-Dios<sup>52</sup> a través del análisis de dos instrumentos: “Effort-Reward Imbalance questionnaire” (ERI) y el “New Job Demand Scale”. Por otro lado, 13 artículos respetaron de cinco o seis de estas recomendaciones mientras que 5 artículos lo hicieron con tres o menos fases. Sólo un artículo incumplió todas recomendaciones<sup>28</sup>.

De las recomendaciones más respetadas, tenemos la de justificación del estudio (19 artículos) seguida del análisis cualitativo de los ítems (18 artículos) o la delimitación del constructo (16 artículos) Las más incompletas fueron el análisis estadístico de los ítems (9 artículos) y la validez de constructo (13 artículos). Las más infringida fue la validez de constructo (4 artículos).

Al centrar el examen en los artículos que aplicaron y describieron los cuatro criterios básicos de análisis psicométrico (validez de constructo, validez externa o convergente-discriminante y fiabilidad por test retest y consistencia interna), sin incluir la rigurosidad de las fases 5,6 y 7 de Carretero, el resultado son 11 artículos<sup>23,24,26,27,29,32-36,42</sup>. En otras palabras, más del 50% de artículos describieron la presencia de este tipo de validaciones aunque fuera de forma

incompleta o dudosa. De todas ellas, la más aplicada fue la validez externa de tipo convergente-discriminante y la menos respetada fue la validez interna o de constructo. Los instrumentos de medición reflejados fueron el ERI (en 4 ocasiones), la NSS (2 ocasiones), el JCQ (1 ocasión) y los 4 instrumentos de nueva creación “Medical Personnel Stress Survey” (MPSS-R) “The Psychiatric Nurse Job Stressor Scale” (PNJSS), la “Stressor Scale for Clinical Research Coordinators” (SSCRC) y “Work Stress Inventory for nurses in Oncology”. Respecto a la validez de constructo, referida al grado en que cada prueba refleja el constructo que dice medir, fue estudiada a través de dos modelos factoriales: análisis factorial exploratorio (EFA) y análisis factorial confirmatorio (CFA)<sup>54</sup>. La mayoría de estudios sólo usaron el EFA con método de extracción por análisis de componentes principales (PCA) y rotación de varimax o promax. Otros dos estudios<sup>38,40</sup> usaron sólo el CFA usando índices comunes como el índice de ajuste comparativo (CFI), error de aproximación por RMSEA, prueba de bondad de ajuste (AGFI) y la media cuadrática residual (RMR). Dos estudios<sup>32,35</sup> usaron tanto EFA como CFA.

Otro subnivel de la validez de constructo sería la validez convergente-discriminante. La primera está basada en el estudio de las relaciones entre el test y otros tests que supuestamente miden lo mismo o con otros constructos con los que tendría que mostrar relación (validez convergente). La segunda mide otras variables teóricamente relevantes y de las que debería diferenciarse (validez discriminante)<sup>53</sup>. Alrededor de 13 artículos exploraron esta validez externa a través de la correlación con otras escalas y/o constructos semejantes o contrarios<sup>24-27,30,32-36,39</sup>. De todos ellos, siete describieron tanto correlación negativa como positiva entre las escalas estudiadas y sus constructos<sup>24,26,27,30,32,36,39</sup>. Las escalas y constructos más utilizados para su convergencia y discriminancia fueron los siguientes: para el concepto de salud general, las escalas “The General health Questionnaire” (GHQ), “The Chinese Health Questionnaire”(CHQ-12) y el SF-36. Para síndrome de burnout escalas como “Maslach Burnout Inventory” (MBI), “Maslach Burnout Inventory-General Survey” (MBI-GS) y la escala de efectos psíquicos del burnout. La satisfacción laboral se correlacionó a través de escalas como “The Work Subscale of the Job Description Index” o el “Job Satisfaction Questionnaire” (JSQ), la ansiedad por “IPAT Anxiety Scale” y “The Affect Rating Scale” e incluso otros estresores laborales y sus efectos a través del “Brief Job Stress Questionnaire” (BJSQ). Finalmente, otros artículos<sup>29,42</sup> valoraron únicamente la validez discriminante y 6 artículos<sup>28,31,37,38,40,41</sup> no valoraron ninguna de las dos.

Si hablamos de fiabilidad y consistencia interna, sólo 13 artículos aplicaron fiabilidad por test-retest junto con consistencia interna<sup>23,25-27,29,31-33,35,36,41,42</sup>. Por otro lado, 8 artículos aplicaron únicamente consistencia interna. De los 21 artículos que mostraron esta consistencia interna, el 100% centraron en el Alfa de Cronbach con valores mínimos de 0,48 y valores máximos de 0,92. Todos ellos tuvieron Alfas de Cronbach >0,70 exceptuando casos puntuales como con el instrumento MPSS-R<sup>34</sup>, para los factores "actitudes negativas del paciente" y "Distrés somático". Otros casos fueron <sup>25</sup> el Post-Code Nurse Stress, para el factor 4 de "Sentirse quemado" y <sup>42</sup> con el cuestionario ERI para el factor "esfuerzo". En relación a los test.retest, la mayoría mostraron correlaciones >0,40 exceptuando para el "Stres of consiente Questionnaire"<sup>41</sup> y la herramienta de medición de burnout y estrés <sup>31</sup>.

#### Otros datos de la calidad psicométrica.

A nivel de cumplimiento de las fases de construcción y evaluación cualitativa de los ítems, la mayoría de estudios (18 en total) siguieron buenos criterios de descripción, eliminación, creación y traducción de los ítems de la escala estudiada.

En relación con el análisis estadístico de los ítems 11 cumplieron con las recomendaciones, 9 lo hicieron de forma incompleta y sólo 1 no analizó los ítems a nivel estadístico.

### **Discusión**

El objetivo principal de esta revisión fue explorar los distintos instrumentos de medición del estrés psicológico y laboral en personal de enfermería. De los artículos revisados se extrajo una extensa cantidad de instrumentos de los cuales destacaron el "Effort-Reward Imbalance Questionnaire" (ERI), el "Job Content Questionnaire" (JCQ), la "Nursing stress scale" (NSS) y "The Stress of Conscience Questionnaire" (SCQ), por ser los más frecuentes en su aplicación en muestras enfermeras. Sin embargo, estos instrumentos sólo correspondieron a la mitad de la literatura seleccionada, siendo el resto de escalas más variables pero poco replicadas. De igual manera, la globalidad de los resultados obtenidos en esta revisión, pone en evidencia una gran heterogeneidad en los instrumentos de medición de estrés, estrechamente ligada a la gran versatilidad del constructo estudiado y la complejidad de la profesión enfermera. Tal como se desprendía de la justificación de esta revisión, estas dos

peculiaridades fueron los motivos principales para explorar los instrumentos de medición del estrés en personal de enfermería.

Según Carretero<sup>52</sup>, uno de los primeros pasos para desarrollar cualquier test psicométrico es realizar una rigurosa descripción del constructo a través de una revisión bibliográfica previa, definición de componentes y relaciones y bajo modelos teóricos de referencia. En este sentido, los resultados mostraron ejemplos tan clásicos como las escalas ERI y JCQ cuyas características más homogéneas las han hecho resaltar en la literatura de las últimas décadas (referencias). Sin embargo, estos mismos instrumentos fueron inicialmente creados para el estrés ocupacional en grandes sectores laborales clasificados por ocupaciones (por ejemplo, “blue collar” o trabajadores manuales y “White collar” o trabajadores de oficina)<sup>55</sup>. Por lo tanto, no es de extrañar encontrar una tendencia a ser aplicados de forma muy global en grandes hospitales y con enfermeras generalistas y polivalentes<sup>56</sup>. A esto se le añade la falta de concreción de las unidades asistenciales en las que ejercían los profesionales de enfermería, hecho que tiende a menguar la singularidad y especificidad de cada servicio y sus situaciones de estrés. Por consiguiente, a pesar de que el ambiente hospitalario ha sido definido como uno de los más estresantes a nivel sanitario<sup>57</sup>, se ha echado en falta no sólo la aplicación de estas escalas (ERI y JCQ) en enfermeras especialistas, sino también en ambientes extrahospitalarios. Por ejemplo centros residenciales, centros de psiquiatría, unidades de emergencias, unidades de cuidados intensivos o servicios de oncología entre otros. Específicamente, estas áreas han sido descritas en la literatura como algunas de las más estresantes del ámbito enfermero<sup>58-60</sup>, mientras que las residencias de ancianos tienen un alto componente de estrés ético o angustia moral<sup>61</sup>.

Hubo algunas excepciones al problema anteriormente planteado: el test “Post-Code Stress Scale”<sup>25</sup> y el “New Job Demand Scale” que, bajo modelos teóricos concretos estudiaron dos muestras de enfermeras de cuidados intensivos y emergencias respectivamente. Y también el estudio de Weyers<sup>40</sup> con el instrumento de ERI, que analizó una parte de muestra hospitalaria y otra parte de atención primaria, aunque con poca diferenciación de ambas a la hora de aplicar la escala.

Desde otras perspectivas de definición del constructo no asociadas a modelos teóricos concretos, la visión del estrés como estímulo (estresores) fue una de las más usadas en esta revisión. En concreto, la gran mayoría de instrumentos originales únicamente creados para personal de enfermería, se apoyaron en esta perspectiva y se aplicaron en ámbitos más especializados de la profesión como la oncología (escala WSINO), los coordinadores de

investigación (escala SSCRC), la psiquiatría (PNJSS) o la geriatría (SDCS). La adaptación de la “Nursing Stress Scale”<sup>27</sup> a la cultura china, fue uno de los pocos ejemplos de estudios sobre estresores laborales en ambiente hospitalario que diferenció entre unidades de cuidados intensivos, medicina interna y cirugía. Sin embargo, hay que señalar que a pesar de que la NSS esté mayoritariamente enmarcada en estímulos estresantes, la redacción de algunas situaciones parecería orientarse hacia el estrés como respuesta (“sentirme inadecuadamente preparado”, “miedo de cometer errores”...). El único instrumento que valora el estrés desde su perspectiva de respuesta (“Medical Personnel Stress Survey”<sup>34</sup>) lo hizo también basándose en ambientes de trabajo más concretos (urgencias/emergencias). Bajo la perspectiva del estrés como interacción entre ambiente y sujeto estuvo la “Stress of Conscience Questionnaire” en enfermeras de hospital y en personal de geriatría<sup>30</sup> con una gran rigurosidad en la descripción del constructo del estrés moral o ético debido a las particularidades de este ambiente laboral. Sin embargo, a pesar de que la definición del constructo, marcos teóricos y muestras de aplicación de los instrumentos son pasos indispensables para el desarrollo de test psicométricos, es en su validez y fiabilidad donde realmente radica la funcionalidad de lo que miden.

El objetivo secundario de esta revisión fue conocer el tipo de validaciones psicométricas más usadas en el estudio de los instrumentos de medición de estrés en personal de enfermería. También se incluyó la evaluación de su calidad a nivel metodológico a través de varios autores especialistas en psicometría<sup>50-54</sup> se extrajeron una serie de datos de interés. A nivel de calidad metodológica y psicométrica, destaca el hecho de que los artículos que cumplieron con las 7 recomendaciones establecidas por Carretero-Dios y Pérez (tablas 3 y 4) estuvieran dedicados a dos de los instrumentos bajo un modelo conceptual bien establecido: instrumento Effort–Reward Imbalance Questionnaire o ERI (el más reproducido en la revisión) y la “New Job Demand Scale”<sup>36</sup>.

De los instrumentos que cumplieron con los cuatro criterios básicos de validación psicométrica (validez de constructo, convergente-discriminante y fiabilidad por test retest y consistencia interna) ya fuera de forma completa o no, predominan nuevamente escalas repetidas en esta revisión: la “Nursing Stress Scale” y el “Job Content Questionnaire”. También se incluyeron dos instrumentos de única aplicación, desarrollados y creados para personal de enfermería “The Psychiatric Nurse Job Stressor Scale” (PNJSS), y la “Stressor Scale for Clinical Research Coordinators” (SSCRC). Cabe destacar un hecho en este apartado de criterios básicos de validación. En los resultados descritos en esta revisión, se puede

comprobar que el tipo de validez menos respetada por los estudios encontrados es la validez de constructo o validez interna. Recordemos que ésta validez es el grado en que cada test refleja el constructo que dice medir (en este caso el estrés) y que es valorada como la única admisible en la literatura de las últimas décadas<sup>54</sup>. Se podría llegar a afirmar según los datos que aporta Carretero en sus recomendaciones<sup>52</sup>, que sin validez de constructo, el instrumento no tiene bases sólidas para llevar a cabo los objetivos de estudio y justificación. Y sin embargo, la validez convergente-discriminante y la fiabilidad y consistencia internas son mucho más respetadas y descritas para cualquiera de los instrumentos.

Volviendo a los datos descritos sobre los instrumentos con una base conceptual bien arraigada, se podría concluir, de forma prematura, que estas escalas más reproducidas son también las que tienen mayor validación psicométrica (JCQ, ERI, NSS, New Job Demand Scale). Sin embargo, se debe profundizar en la calidad para determinar las posibles discrepancias con esta hipótesis.

Si centramos la exploración en el análisis de la validez de constructo, que ya hemos descrito como una de las bases para desarrollar cualquier test psicométrico<sup>52,54</sup>, se destaca el grado de rigor o calidad de los instrumentos. Como ya habíamos planteado en otros apartados, el análisis factorial (tanto exploratorio como confirmatorio) es el método de elección para la validez de constructo en todos los artículos revisados. Como guía para evaluar el análisis factorial exploratorio (EFA) se siguieron las recomendaciones descritas en el anexo 2. Es decir, se valoraron aquellos estudios que siguieran la medida de la adecuación muestral previa de Kaiser-Meyer-Olkin junto con la prueba de esfericidad de Bartlett, método de extracción por componentes principales (PCA) y rotación tipo varimax. También debían describir el número de factores, las saturaciones de los ítems y la proporción de la varianza explicada. Se valoró igualmente, que contaran con un análisis factorial confirmatorio (CFA) de refuerzo. Como guía para los resultados del CFA se siguieron los criterios de Kelloway<sup>62</sup> por los que se considera que el índice de ajuste comparativo (CFI) debe estar en torno a 0.95 para que el modelo se ajuste adecuadamente a los datos. El error de aproximación por RMSEA  $< 0.05$  indica un buen ajuste a los datos si, además, el intervalo de confianza al 90% (I.C.) se sitúa entre 0 y 0.05. La Prueba de bondad de ajuste (AGFI)  $> 0,9$  con al menos un 90% de la varianza explicada por el modelo. La media cuadrática residual (RMR) entre 0-0,10 es aceptable.

Los resultados de la revisión en base a la calidad de la validez de constructo (apoyo en datos de tabla del anexo 1) evidencian un bajo cumplimiento de todos los criterios de EFA y CFA descritos. Los de mejor calidad fueron los artículos que cumplieron con las 7 recomendaciones establecidas por Carretero-Dios y Pérez de forma completa. Es decir, por un lado Msauel con la escala ERI a través de EFA con método de extracción por PCA y rotación varimax KMO (0,84), esfericidad de Bartlett ( $<0,001$ ), peso factorial  $>0,30$  y comunalidad o proporción de varianza explicada del 41,19 % del modelo de la escala aplicada. Y por otro, Sundin<sup>36</sup> con la “New Job Stress Scale” (KMO= 0,87/ Prueba de esfericidad de Bartlett:  $p<0,0005/PA>1,201$ , scree test y varianza explicada de casi el 70%) aunque sin apoyo del CFA. Pero otros instrumentos también demostraron valores óptimos en su validez de constructo. Por ejemplo Yada con el instrumento “The Psychiatric Nurse Job Stressor Scale”<sup>32</sup>, (KMO= 0,788//Esfericidad de Bartlett  $<0,001$  y scree-test sin análisis paralelo. Peso factorial  $>0,40$ , varianza explicada del 51% y CFA  $X^2/d.f.$  (343.189/196) ratio = 1.750( $P < 0.01$ ), GFI = 0.910, AGFI = 0.883, CFI = 0.924, RMSEA = 0.050). También Matsumoto<sup>35</sup> con la “Stressor Scale for Clinical Research Coordinators” (SSCRC) que se apoyó no sólo en el EFA sino también en la aplicación de un análisis factorial confirmatorio posterior. Igualmente, describió haber realizado índice de análisis paralelo, KMO y Esfericidad de Bartlett pero los resultados de los mismos no sale descritos en el estudio. Únicamente se describen los siguientes: peso factorial  $>0,40$  y CFA con los índices de análisis al límite de lo considerado aceptable: GFI = 0,910, AGFI = 0,889, CFI = 0,922, RMSEA = 0,054).

Con esto, encontramos que no todos los instrumentos planteados anteriormente como válidos, alcanzan niveles de exigencia aceptables. Pero, por el contrario, es cierto que todos aquellos estudios que sólo presentan uno o dos tipos de validez (sin contar la de constructo) o la presentan de forma muy incompleta, no tienen una metodología en su análisis factorial de rigor o de alta calidad. Por lo tanto, se admite que para las escalas SSCRC, PNJSS y “New Job Demand Scale” se cumple esa calidad en la validez psicométrica de constructo y se podría admitir que parten de un rigor más elevado que el resto a nivel metodológico Sin embargo, a pesar de que los tests ERI y JCQ fueran los más reproducidos y estudiados, sólo una de sus adaptaciones a enfermería demostró una validez aceptable<sup>33,42</sup>.

Tras la validez de constructo, se valoraría si estos instrumentos que se presentan inicialmente como de gran calidad metodológica, cumplen con el resto de validaciones.

El estudio de la validez convergente-discriminante, que en algunos ámbitos de la psicometría se considera parte de la de constructo, supone el soporte a partir del cual interpretar las puntuaciones de la herramienta de evaluación y, por lo tanto, el modo de otorgar significado psicológico a un dato numérico<sup>63</sup>. En relación a los resultados expuestos, se han hallado correlaciones tanto positivas como negativas para explicar la varianza convergente-discriminante en varios instrumentos. Sin embargo, en su análisis se ha esclarecido o bien una falta de significación estadística ( $p > 0,05$ ) o bien bajas correlaciones. Los ejemplos más claros de correlaciones positivas elevadas y estadísticamente significativas ( $p < 0,01$ ) se encontraron en la evaluación del instrumento WSINO para enfermeras de oncología con la dimensión de agotamiento emocional (intervalo  $r = 0,34-0,64$ ) de la escala MBI del burnout. También se observaron datos fiables en la escala SCQ (Stress of Conscience) para enfermeras de geriatría, que se correlacionó de forma negativa con la satisfacción laboral a través del “Job Satisfaction Questionnaire” (JSQ) ( $r = -0,45$   $p < 0,01$ ) y con “The Strain in Dementia Care Scale” (SDCS) en correlación positiva ( $r = 0,61$   $p < 0,01$ ). También el instrumento para estrés en enfermería “The Psychiatric Nurse Job Stressor Scale” (PNJSS) en enfermeras de salud mental, estuvo correlacionado con la subescala del “Stress reaction’ scale” del “Brief Job Stress Questionnaire” (BJSQ). ( $r = 0,45$   $p < 0,05$ ). La escala SSCRC se relacionó de forma positiva con los factores agotamiento, cinismo, K6 ( $> 0,25$ ) bajo significancia estadística ( $p < 0,01$ ).

En lo que respecta a las escalas con buena validez de constructo, tenemos que Msasouel utiliza para la ERI una validación discriminante en la que compara la misma escala para distintos grupos de la misma muestra de enfermeras (comparación de medias Kruskal-wallis y Mann-Whitney). Sus valores no fueron estadísticamente significativos para sexo, edad o turnos laborales.

Nuevamente, se extrae de los resultados que las escalas anteriormente mencionadas como de alta calidad, a pesar de no cumplir con todas las fases de las tablas 3 y 4, tienen validaciones óptimas y de calidad metodológica.

El último nivel descrito en esta revisión es el de fiabilidad y consistencia interna. Un instrumento de medida es fiable cuando proporciona puntuaciones estables y consistentes. Así, mientras la validez se refiere a la coincidencia de dos medidas de algún objeto mediante métodos diferentes, la fiabilidad se refiere a la coincidencia de la medida de algo con métodos iguales o muy similares. Para ello se utilizan, a gran escala, dos técnicas: la de pruebas repetidas o test-retest, que permite determinar la estabilidad de una medida en el tiempo; y el coeficiente alfa de Cronbach que mide la consistencia interna. En el anexo 2 se desarrollan de

forma más precisa las recomendaciones sobre fiabilidad y consistencia interna entre las que cabe destacar el uso del Alfa de Cronbach para escalas con menos de 30 ítems tipo Likert (valores entre 0,70-0,80 óptimos) y el uso de la correlación media inter-item para escalas con más de 30 ítems (valores entre 0,15 y 0,50 aceptables).

Como se puede comprobar, únicamente 8 artículos realizaron fiabilidad por test-retest además de la consistencia interna. Sin embargo, hubo un escaso cumplimiento de los valores aceptables de correlación del test-retest ( $0.80 < r < 0.90$   $-0.90 < r < -0.80$ ). Uno de los únicos instrumentos que tuvo una correlación adecuada fue el Job Content Questionnaire del estudio de Escribá (2001) que, además, tuvo una consistencia interna por Alfa de Cronbach (<30 ítems) entre 0,74 y 0,87, valores considerados óptimos. Sin embargo, a nivel de validez de constructo, este instrumento no alcanza la calidad de los anteriormente citados.

Por otro lado, el “Stress of Conscience Questionnaire”, tuvo correlaciones límite (<0,80) para el estudio de Glasberg<sup>41</sup>, llegando a ser de 0,85 en el posterior estudio de Orrung<sup>30</sup>. Los valores de Alfa para ambos superaron 0,80, no siendo del todo aceptables según las recomendaciones. La “Post-Code Stress Scale” se aproximó a la correlación de test-retest óptima sin llegar a superar los 0,80, pero teniendo valores de Alfa entre 0,57-0,77.

Otro dato de relevancia relacionado con la consistencia interna, es el cumplimiento de la aplicación de correlaciones inter-item para tests superiores a los 30-40 ítems. Los estudios del instrumento CNSQ<sup>22</sup> (50 ítems), “New tool for burnout and stress” (53 ítems) y NSC<sup>37</sup> (43 ítems) lo aplican pero las dos últimas escalas no describen las puntuaciones de la correlación. Y por otro lado, la escala WSINO, una de las más completas a nivel de validación psicométrica, no sólo no tiene test-retest sino que, con 50 ítems, sólo hace uso del Alfa de Cronbach para su consistencia.

Todos los instrumentos de esta revisión mostraron valores de Alfa de Cronbach >0,70 exceptuando casos puntuales como el instrumento MPSS-R<sup>34</sup>, para los factores "actitudes negativas del paciente" y "Distrés somático". Otros casos fueron <sup>25</sup> con el Post-Code Nurse Stress, para el factor 4 de “Sentirse quemado” y <sup>42</sup> con el cuestionario ERI para el factor “esfuerzo”.

De los 21 artículos que mostraron consistencia interna la mayoría se centraron en el alfa de Cronbach por dimensiones o subescalas o en el total de la escala estudiada, sin tener en cuenta el alfa de los ítems descritos.

Tras la revisión de los resultados hallados podemos extraer que los mejores instrumentos a nivel de reproductibilidad, definición de constructo, validación y fiabilidad son: “The

Psychiatric Nurse Job Stressor Scale” (PNJSS) de Yada, el “New Job Stress Scale” de Sundin, la “Escala Effort-Reward Imbalance” de Msaouel y, a pesar de que no tenga la descripción de la validez de constructo completa, el Job Content Questionnaire del estudio de Escribá<sup>33</sup> por ser uno de los pocos estudios que cumplen con más criterios de fiabilidad y consistencia además de tener una validez convergente-discriminante estadísticamente significativa (a pesar de no ser tan elevada).

## **Conclusiones**

Los resultados de esta revisión han aportado evidencia de la gran variabilidad en las escalas, cuestionarios y tests psicométricos dirigidos a medir el estrés en personal de enfermería. Esta cuestión se ha relacionado con la gran versatilidad del constructo estudiado y la complejidad de la profesión enfermera. En este sentido, se ha llegado a la conclusión que, si bien el ambiente hospitalario sigue siendo el origen de altos niveles de estrés en cualquiera de sus formas, la enfermera en otros medios extrahospitalarios también sufre de elevadas cargas laborales y éticas. A esto se le añade que la perspectiva clásica de la enfermera como ente polivalente o generalista está siendo sustituida por la de profesional cada vez más especializada y con mayores niveles de autonomía. Esto plantea nuevas alternativas al análisis del ambiente enfermero, que comienza, poco a poco, a estar enfocado a competencias más concretas y complejas como se ha podido comprobar en esta revisión. En concreto, aún no se han especializado las herramientas como la ERI o la JCQ para personal de enfermería a pesar de contar con validaciones psicométricas aceptables y/o óptimas. En cambio, las nuevas herramientas que exploran las peculiaridades de esta nueva “raza” de enfermeras, pecan todavía de falta de validaciones psicométricas, de dudosa definición y contextualización del constructo y de baja reproductibilidad (o repetición de ensayos) en más muestras de enfermería.

Por otro lado, se quiere recalcar el hecho de que en el desarrollo de instrumentos de medición de personal de enfermería, se hayan dado tan baja calidad en la validez de constructo, siendo ésta de suma importancia para el desarrollo de cualquier test psicométrico. Esta situación podría deberse a dos aspectos de la metodología de esta revisión. La primera es las altas exigencias en los criterios de inclusión y exclusión escogidos para las búsquedas bibliográficas que se centraron exclusivamente en estudios de validación, creación y desarrollo de instrumentos. Y la segunda es la elección de criterios de análisis de la calidad

metodológica tan rigurosos como los expuestos en esta revisión por las 7 fases de Carretero. En el futuro, se podrían realizar nuevas búsquedas que se centren en otros aspectos del estrés secundarios a su validación o incluso a aquellos factores protectores para la enfermera como lo son la capacidad de afrontamiento, la resiliencia, la satisfacción laboral o la valoración social.

## **Bibliografía**

1. Bartlett D. Stress: Perspectives and processes. McGraw-Hill International;1998
2. Sierra JC, Ortega V, Zubeidat I. Ansiedad, angustia y estrés: tres conceptos a diferenciar. Rev mal-estar subj. 2003;3(1):10-59.
3. Sutherland VJ, Cooper CL. Understanding stress: A psychological perspective for health professionals. Chapman & Hall/CRC, 1990.
4. Lazarus RS, Folkman S. Stress, appraisal, and coping. New York, NY: Springer;1984.
5. Karasek R, Theorell T. Healthy work: Stress, productivity and the reconstruction of working life. New York: Basic Books;1990.
6. Siegrist J. Adverse health effects of high effort – low reward conditions at work. J Occup Health Psychol. 1996;1:27-43.
7. Glazer S, Gyurak A. Sources of occupational stress among nurses in five countries. Int J Intercult Relat.2008;32(1):49-66.
8. Sveinsdottir H, Biering P, Ramel A. Occupational stress, job satisfaction, and working environment among Icelandic nurses: a cross-sectional questionnaire survey. Int J Nurs Stud. 2006;43(7):875-889.
9. Chang EM, Daly J, Hancock KM, Bidewell JW, Johnson A, Lambert VA, Lambert CE. The relationships among workplace stressors, coping methods, demographic characteristics, and health in Australian nurses. J Prof Nurs. 2006;22(1):30-8.
10. Lambert VA, Lambert CE, Itano J, Inouye J, Kim S, Kuniviktikul W, et al. Cross-cultural comparison of workplace stressors, ways of coping and demographic characteristics as predictors of physical and mental health among hospital nurses in Japan, Thailand, South Korea and the USA (Hawaii). Int J Nurs Stud. 2004;41(6):671-684.
11. Kérouac S. El pensamiento enfermero. Barcelona: Masson; 2007; 1996.

12. Garrosa E, Moreno-Jiménez B, Liang Y, González JL. The relationship between socio-demographic variables, job stressors, burnout, and hardy personality in nurses: An exploratory study. *Int J Nurs Stud.* 2008;45(3):418-427.
13. ) Cortese CG, Colombo L, Ghislieri C. Determinants of nurses' job satisfaction: the role of work–family conflict, job demand, emotional charge and social support. *J Nurs Manag.* 2010;18(1):35-43.
14. Pla O, Hernández A, Orquín J, Sanz de Galdeano FJ. La enfermería y el estrés laboral: amistades peligrosas. *Enfermería Integral.* 1999;48.
15. Jourdain G, Chênevert D. Job demands–resources, burnout and intention to leave the nursing profession: A questionnaire survey. *Int J Nurs Stud.* 2010;47(6):709-22.
16. Nunnally J, Bernstein Y. *Teoría psicométrica.* México: McGraw-Hill;1995.
17. Morales ML. *Psicometría aplicada.* México: Trillas;1996.
18. Aiken L. *Tests psicológicos de evaluación.* México: Prentice-Hall;1996.
19. Anastasi A. *Los tests psicológicos.* Madrid: Aguilar; 1986.
20. De Kohan NC. *Teorías psicométricas y construcción de tests.* Buenos Aires: Lugar; 1999.
21. Liberati A Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *J Clin Epidemiol.* 2009; 62:e1–e34.
22. Admi H, Moshe-Eilon Y. Stress among charge nurses: tool development and stress measurement. *Nurs Econ.* 2010;28(3):151-158.
23. Bonnetterre V, Liaudy S, Chatellier G, Lang T, de Gaudemaris R. Reliability, validity, and health issues arising from questionnaires used to measure Psychosocial and Organizational Work Factors (POWFs) among hospital nurses: a critical review. *J Nurs Meas.* 2008;16(3):207-230.
24. Borteyrou X, Truchot D, Rascle N. Development and validation of the Work Stressor Inventory for Nurses in Oncology: preliminary findings. *J Adv Nurs.* 2014;70(2):443-453.
25. Cole FL, Slocumb EM, Mastey JM. A measure of critical care nurses' post-code stress. *J Adv Nurs.* 2001;34(3):281-288.

26. Escribà V, Más R, Cárdenas M, Pérez S. Validación de la escala de estresores laborales en personal de enfermería: «the nursing stress scale». *Gac Sanit.* 1999; 13(3):191-200.
27. Lee MH, Holzemer WL, Faucett J. Psychometric evaluation of the Nursing Stress Scale (NSS) among Chinese nurses in Taiwan. *J Nurs Meas.* 2007;15(2):133-144.
28. Li J, Yang W, Liu P, Xu Z, Cho SI. Psychometric evaluation of the Chinese (mainland) version of Job Content Questionnaire: a study in university hospitals. *Ind Health.* 2004;42(2):260-7.
29. Msaouel P, Keramaris NC, Apostolopoulos AP, Syrmos N, Kappos T, Tasoulis A, et al. The effort-reward imbalance questionnaire in Greek: Translation, validation and psychometric properties in health professionals. *J Occup Health.* 2012;54(2):119-30.
30. Orrung Wallin A, Edberg AK, Beck I, Jakobsson U. Psychometric properties concerning four instruments measuring job satisfaction, strain, and stress of conscience in a residential care context. *Arch Gerontol Geriatr.* 2013;57(2):162-171.
31. Skinner V, Lee-White T, Agho K, Harris J. The development of a tool to assess levels of stress and burnout. *Aust J Adv Nurs.* 2007;24(4):8-13.
32. Yada H, Abe H, Funakoshi Y, Omori H, Matsuo H, Ishida Y, Katoh T. Development of the Psychiatric Nurse Job Stressor Scale (PNJSS). *Psychiatry Clin Neurosci.* 2011;65(6):567-575.
33. Escribà-Agüir V, Mas Pons R, Flores Reus E. Validación del Job Content Questionnaire en personal de enfermería hospitalario. *Gac Sanit.* 2001;15(2): 142-149.
34. Li TC, Lee YD, Lee YY, Chen CC, Chi CH. Validation of the Chinese-version MPSS-R for occupational stress among emergency nurses. *J Psychosom Res.* 2001;51(1):379-385.
35. Matsumoto K, Sumino K, Fukahori H, Kitaoka K, Kamibeppu K, Nagamura F. Stressor Scale for Clinical Research Coordinators: development and psychometric testing. *J Adv Nurs.* 2012;68(7):1636-1645.
36. Sundin L, Hochwälder J, Bildt C. A scale for measuring specific job demands within the health care sector: Development and psychometric assessment. *Int J Nurs Stud.* 2008;45(6):914-23.
37. Tsai SL. The development of a nurse stress checklist from English to Chinese version. *Issues Ment Health Nurs.* 1997;18(3):247-257.

38. Griep RH, Rotenberg L, Vasconcellos AG, Landsbergis P, Comaru CM, Alves MG. The psychometric properties of demand-control and effort-reward imbalance scales among Brazilian nurses. *Int Arch Occup Environ Health*. 2009;82(10):1163-72.
39. Ortiz VG. Assessment of Psychosocial Stressors at Work: Psychometric Properties of the Spanish Version of the ERI (Effort-Reward Imbalance) Questionnaire in Colombian Workers. *Revista de Psicología del Trabajo y de las Organizaciones*. 2010;26(2):147-156.
40. Weyers S, Peter R, Boggild H, Jeppesen HJ, Siegrist J. Psychosocial work stress is associated with poor self-rated health in Danish nurses: a test of the effort-reward imbalance model. *Scand J Caring Sci*. 2006;20(1):26-34.
41. Glasberg AL, Eriksson S, Dahlqvist V, Lindahl E, Strandberg G, Söderberg A et al. Development and initial validation of the stress of conscience questionnaire. *Nurs Ethics*. 2006;13(6):633-48.
42. Macías Robles MD, Fernández-López JA, Hernández-Mejía R, Cueto-Espinar A, Rancaño I, Siegrist J. Evaluación del estrés laboral en trabajadores de un hospital público español. Estudio de las propiedades psicométricas de la versión española del modelo «Desequilibrio Esfuerzo-Recompensa». *Med Clin (Barc)*. 2003;120(17):652-7.
43. Siegrist J. Adverse health effects of high effort/ low-reward conditions. *J Occup Health Psychol* 1996;1:27-41.
44. Siegrist J, Peter R. The effort-reward imbalance model. *Occup Med* 2000;15:83-7
45. Siegrist J. A theory of occupational stress. In: Dunham J, editor. *Stress in the workplace: Past, present and future*. London: Whurr Publishers; 2000.
46. Karasek R. Job demands, job decision latitude and mental strain: Implications for job redesign. *Adm Sci Q*. 1979; 24:285-306.
47. Karasek R, Theorell T. *Healthy work: Stress, productivity and the reconstruction of working life*. New York: Basic Books; 1990.
48. Karasek R, Brisson C, Kawakami N, Houtman I, Bongers P, Amick B. The Job Content Questionnaire (JCQ): an instrument for internationally comparative assessments of psychosocial job characteristics. *J Occup Health Psychol*. 1998;3:322-55.
49. Biddle BJ. *Role Theory: Expectations, Identities, and Behaviors*. New York: Academic Press; 1979.

50. Cronbach LJ. Fundamentos de la exploración psicológica. Madrid: Biblioteca Nueva;1972.
51. Norbeck, JS. What constitutes a publishable report of instrument development? *Nursing Research*.1985;34(6): 380-382.
52. Dios HC, Meléndez CP. Normas para el desarrollo y revisión de estudios instrumentales. *Int J Clin Health Psychol*. 2005;5(3):521-551.
53. AERA, APA y NCME. Standards for educational and psychological tests. Washington DC: American Psychological Association, American Educational Research Association, National Council on Measurement in Education;1999.
54. Messick S. Standards of validity and the validity of standards in performance assessment. *Educ. meas., issues pract*. 1995;14:5-8.
55. Huang GD, Feuerstein M, Sauter SL. Occupational stress and work-related upper extremity disorders: concepts and models. *Am J Ind Med*. 2002 May;41(5):298-314.
56. Runciman PJ. Marco del CIE para las competencias de la enfermera generalista. Ginebra: CIE, 2003.
57. Tomás J, Fernández L. Estrés laboral en los profesionales de enfermería hospitalarios. *Metas*. 2002;45:6-12.
58. Iglesias MEL, de Bengoa Vallejo RB. Prevalence and relationship between burnout, job satisfaction, stress, and clinical manifestations in Spanish critical care nurses. *Dimens Crit Care Nurs*. 2013;32(3):130-7.
59. Arteaga Ruiz. A. La enfermera ante el síndrome de burnout: estudio comparativo atención primaria-atención especializad. Catálogo de Tesinas de la Biblioteca Nacional de Ciencias de la Salud. Madrid. Escuela Nacional de Sanidad, 1997.
60. Escribà-Agüir V, Martín-Baena D, Pérez-Hoyos S. Psychosocial work environment and burnout among emergency medical and nursing staff. *Int Arch Occup Environ Health*. 2006;80(2):127-33.
61. Källemark S, Höglund AT, Hansson MG, Westerholm P, Arnetz B. Living with conflicts-ethical dilemmas and moral distress in the health care system. *Soc Sci Med*. 2004;58(6):1075-84.
62. Kelloway EK. Using LISREL for structural equation modeling. Thousand Oaks, CA:Sage;1998.
63. Paz MD. Validez. En J. Muñiz (Ed.), *Psicometría* (pp. 499-103). Madrid: Universitat. Pelechano, V. (1997). Prólogo. En G. Buela-Casal y J.C. Sierra (dirs.), *Manual de*

evaluación psicológica. Fundamentos, técnicas y aplicaciones (pp. 31-35). Madrid: Siglo XXI: 1996

## **Anexos**

### **ANEXO 1: Documento adjunto**

### **ANEXO 2: Desarrollo y explicación de las 7 fases de Carretero-Dios<sup>52</sup> y su uso en la revisión.**

Para esta revisión se desarrolló una serie de preguntas y criterios que guían a la hora de conocer el grado de cumplimiento de estas 7 fases de creación y desarrollo de los instrumentos y estudios analizados.

- Fase 1. Descripción de la justificación del estudio: ¿Presenta el estudio información coherente y relevante sobre lo que aportaría el instrumento con respecto a medidas ya existentes?
- Fase 2. Delimitación conceptual del constructo a evaluar: ¿Presenta el estudio una adecuada conceptualización del constructo que incluya definición semántica, relaciones esperadas y/o marco teórico?
- Fase 3. Construcción y evaluación cualitativa de ítems: ¿Se especifica el proceso de creación, selección y eliminación de ítems? ¿Se describe su medición y cantidad total? ¿Representan al constructo estudiado?
- Fase 4. Análisis estadístico de los ítems: ¿Se aplica la batería de ítems a una muestra de participantes semejante a la de la población objetivo en modo validación cruzada? ¿Se presentan alguno de los siguientes cálculos estadísticos: desviación típica y puntuaciones medias, coeficiente de correlación ítem-total corregido y correlación media inter-item, consistencia interna de ítems?
- Fase 5. Estudio de la dimensionalidad del instrumento (estructura interna): ¿Se presenta la validez de constructo de la escala incluyendo los siguientes apartados?
  - Descripción de: estimadores que aseguren la matriz de correlaciones apropiada:
    - Kaiser-Meyer-Olkin (coeficiente KMO >0.7 indica alta correlación y, por tanto, conveniencia de un análisis factorial).

- Prueba de esfericidad de Bartlett (próximo al valor 0,000)
- Métodos de extracción y rotación de factores: recomendación de usar análisis factorial exploratorio con método de extracción por componentes principales y rotación varimax.
- Número de factores resultantes
- Saturaciones o peso factorial de los ítems: señalarse  $\geq 0,40$  o de 0,25-0,30 para cuando las muestras están formadas por más de 300 participantes.
- Cantidad de varianza explicada de la escala total y de cada factor y proporción de la varianza del ítem que es explicada por los componentes principales.
- Fase 6. Estimación de la fiabilidad: ¿Se presentan datos de fiabilidad: test-retest, correlación inter-ítem, consistencia interna siguiendo las siguientes recomendaciones?
  - Alfa de Cronbach para escalas con menos de 30 ítems tipo Likert. Valores entre 0,70-0,80.
  - Correlación media inter-ítem para escalas con más de 30 ítems. Valores entre 0,15 y 0,50.
  - Otro tipo de correlaciones ítem-escala o ítem-factores.
  - Coeficiente de fiabilidad test-retest a través de correlación de Pearson o de Spearman. Correlaciones aceptables  $0,80 < r < 0,90$   $-0,90 < r < -0,80$

En todo caso, las correlaciones deben ser estadísticamente significativas ( $p < 0,05$ ).

- Fase 7. Obtención de evidencias externas de validez: ¿Se presentan datos de validación convergente o discriminante?

De esta forma, estos estudios supondrían el soporte a partir del cual interpretar las puntuaciones de la herramienta de evaluación y, por lo tanto, el modo de otorgar significado psicológico a un dato numérico (Paz, 1996).

#### EJEMPLOS DE CUMPLIMIENTO INCOMPLETO O DUDOSO DE LAS FASES

Fase 4: Sólo 1 cálculo estadístico sin realización de un test en una muestra de tipo “piloto”.

Fase 5: Sólo descritos o realizados métodos de extracción y rotación de factores, número de factores resultantes, saturaciones o peso factorial de los ítems y cantidad de varianza explicada para cada factor.

Fase 6: Sin aplicación de fiabilidad por test-retest.